



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - KS141501
ANALISIS FAKTOR- FAKTOR YANG MEMPENGARUHI
KESEMBUHAN DAN KEMATIAN PADA PASIEN STROKE
DENGAN MENGGUNAKAN ASSOCIATION
MINING

ERINA SISKA DEWI
NRP 5210 100 151

Dosen Pembimbing
Wiwik Anggraeni, S.Si, M.Kom.
Renny Pradina Kusumawardani, S.T., M.T.

JURUSAN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2015



ITS
Institut
Teknologi
Sepuluh Nopember

FINAL PROJECT - KS141501
**ANALYSIS OF FACTORS AFFECTING THE HEALING
AND MORTALITY IN STROKE PATIENTS WITH
ASSOCIATION RULE MINING**

ERINA SISKA DEWI
NRP 5210 100 151

Supervisors
Wiwik Anggraeni, S.Si, M.Kom.
Renny Pradina Kusumawardani, S.T., M.T.

DEPARTMENT OF INFORMATION SYSTEMS
Faculty of Information Technology
Institut Teknologi Sepuluh Nopember
Surabaya 2015

**ANALISIS FAKTOR – FAKTOR YANG
MEMPENGARUHI KESEMBUHAN DAN
KEMATIAN PADA PASIEN STROKE DENGAN
MENGUNAKAN ASSOCIATION RULE MINING**

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh:

ERINA SISKA DEWI
NRP. 5210 100 151

Surabaya, 2 Januari 2015

Ketua Jurusan Sistem Informasi

Dr. Eng. Febrilivan Samopa, S.Kom., M.Kom.
NIP. 19730219 199802 1 001

**ANALISIS FAKTOR – FAKTOR YANG
MEMPENGARUHI KESEMBUHAN DAN
KEMATIAN PADA PASIEN STROKE DENGAN
MENGUNAKAN ASSOCIATION RULE MINING**

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh :

ERINA SISKA DEWI
NRP. 5210 100 151

Disetujui Tim Penguji : Tanggal Ujian : 16 Januari 2015
Periode Wisuda : Maret 2015

Wiwik Anggraeni, S.Kom, M.Kom.


(Pembimbing I)

Renny Pradina Kusumawardani, S.T, MT.


(Pembimbing II)

Erma Suryani, ST, MT, Ph.D.


(Penguji I)

Mahendrawati ER., ST., M.Sc., Ph.D.


(Penguji II)

ANALISIS FAKTOR – FAKTOR YANG MEMPENGARUHI KESEMBUHAN DAN KEMATIAN PADA PASIEN STROKE DENGAN MENGUNAKAN ASSOCIATION RULE MINING

Nama Mahasiswa : Erina Siska Dewi
NRP : 5210 100 151
Jurusan : Sistem Informasi FTIf-ITS
Dosen Pembimbing : Wiwik Anggraeni, S.Si, M.Kom
Renny Pradina Kusumawardani, S.T, MT

Abstrak

Penyakit stroke mengalami peningkatan jumlah penderita dan resiko kematian yang meningkat setiap tahun, oleh karena itu diperlukan penanganan medis yang tepat pada penderita stroke. Selain mempertimbangkan faktor resiko yang mempengaruhi seseorang sebelum terjangkit stroke, perlu diketahui faktor – faktor yang mempengaruhi kesembuhan dan kematian dari penderita stroke pada saat menjalani perawatan. Pada penelitian ini akan dianalisis faktor – faktor yang mempengaruhi kesembuhan dan kematian pasien stroke selama 14 hari perawatan dan 6 bulan pemantauan. Karena banyaknya faktor yang terlibat, maka untuk mengetahui faktor mana yang benar – benar memiliki pengaruh pada kesembuhan dan kematian pasien stroke, akan dilakukan analisis dengan menggunakan association rule mining dan algoritma apriori.

Data yang digunakan bersumber dari database International Stroke Trial , data akan dibagi menjadi dua bagian yaitu data diagnosis awal dan data survival pasien. Dari masing – masing data tersebut akan dibagi menjadi data training dan testing, dengan proporsi 70% data training dan 30% data testing. Keluaran yang dihasilkan dari penggalan di dalam data training berupa rule yang memenuhi batas ambang minimum support dan minimum confidence. Nilai minimum support ditentukan berdasarkan proporsi data, sedangkan nilai minimum confidence

ditentukan minimal 0.6 dari proporsi kelas dalam data. Kualitas rule yang dihasilkan diukur dengan menggunakan perhitungan support, confidence, lift dan coverage. Setelah penggalian rule dari data training selesai, rule – rule tersebut akan diuji ke data testing dan diukur akurasi.

Hasil penelitian yang diperoleh adaah berupa rule yang menggambarkan pola, pola tersebut dapat digali untuk menemukan faktor – faktor yang mempengaruhi diagnosis penyakit stroke dan hasil akhir dari kondisi pasien (meninggal/cacat/sembuh/belum sembuh). Rata – rata akurasi yang didapatkan sebesar 49,3% untuk data diagnosis dan 96% untuk data kelangsungan hidup pasien.

Kata Kunci: Association rule mining, algoritma apriori, stroke.

ANALYSIS OF FACTORS AFFECTING THE HEALING AND MORTALITY IN STROKE PATIENTS WITH ASSOCIATION RULE MINING

Name : Erina Siska Dewi
NRP : 5210 100 151
Departement : Information System FTIF-ITS
Supervisor : Wiwik Anggraeni S.Si, M.Kom
Renny Pradina Kusumawardani, S.T, M.T

Abstract

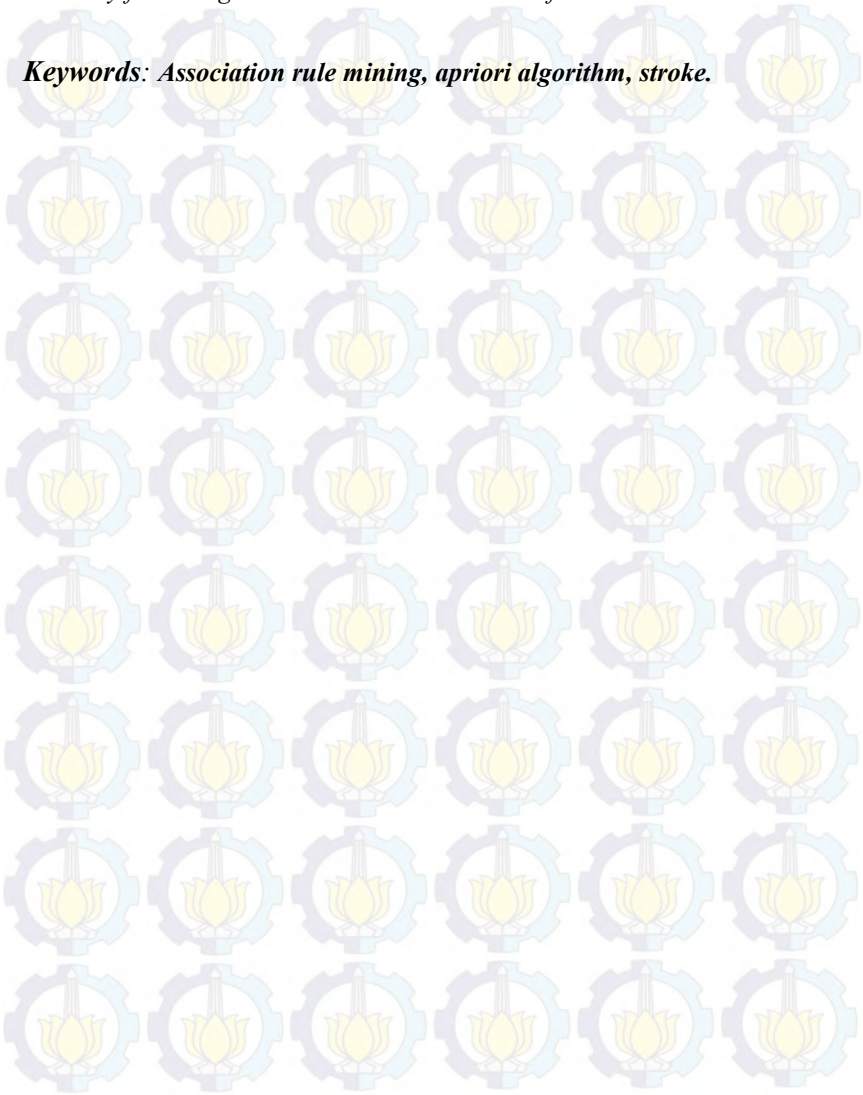
Stroke has increased the number of patients and the risk of death every year, therefore stroke patients need proper medical treatment. Besides considering the factors that affect a person's risk before contracting a stroke, we have to be aware of factors that affect healing and mortality of stroke patients during treatment. In this study, will be analyzed factors that affect healing and mortality of stroke patients during 14 days of treatment and 6 months of monitoring. Because there are a lot of factors involved, then to determine which factors that have an influence or affecting the healing and mortality of stroke patients, the data will be analyzed by using association rule mining and the apriori algorithm.

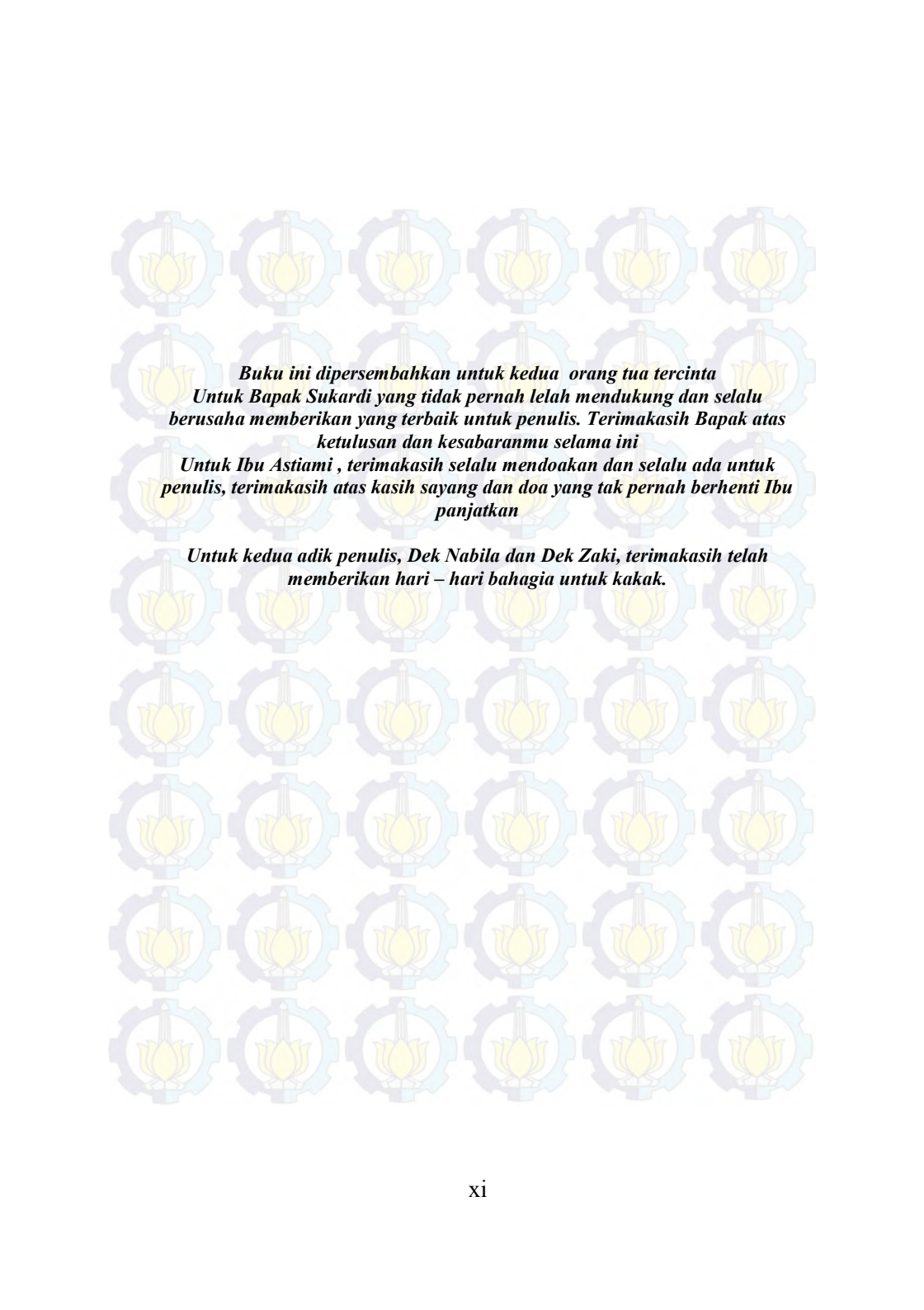
Data is obtained from the database of the International Stroke Trial, it will be divided into two parts, the diagnosis data and patient survival data. Each of them will be divided into training data and testing, with a proportion of 70% training data and 30% testing data. The output of rule mining in the training data are rules or patterns that meets the minimum support threshold and minimum confidence. Minimum support value is determined based on the proportion of the data, while minimum confidence value is 0.6 from each class proportion. The quality of the output is measured by using support, confidence, lift and coverage. After rules extraction from training data is completed, the rules will be tested to the testing data and accuracy value for each rule will be measured.

The results obtained was largely in the form of rules that describe the pattern, the pattern can be explored to find factors that affect the diagnosis of stroke and the final patiens condition

(Dead/Dependent/Recovered/ not recovered). The average value of accuracy from diagnosis data is 49.3% and 96% for the survival data.

Keywords: Association rule mining, apriori algorithm, stroke.



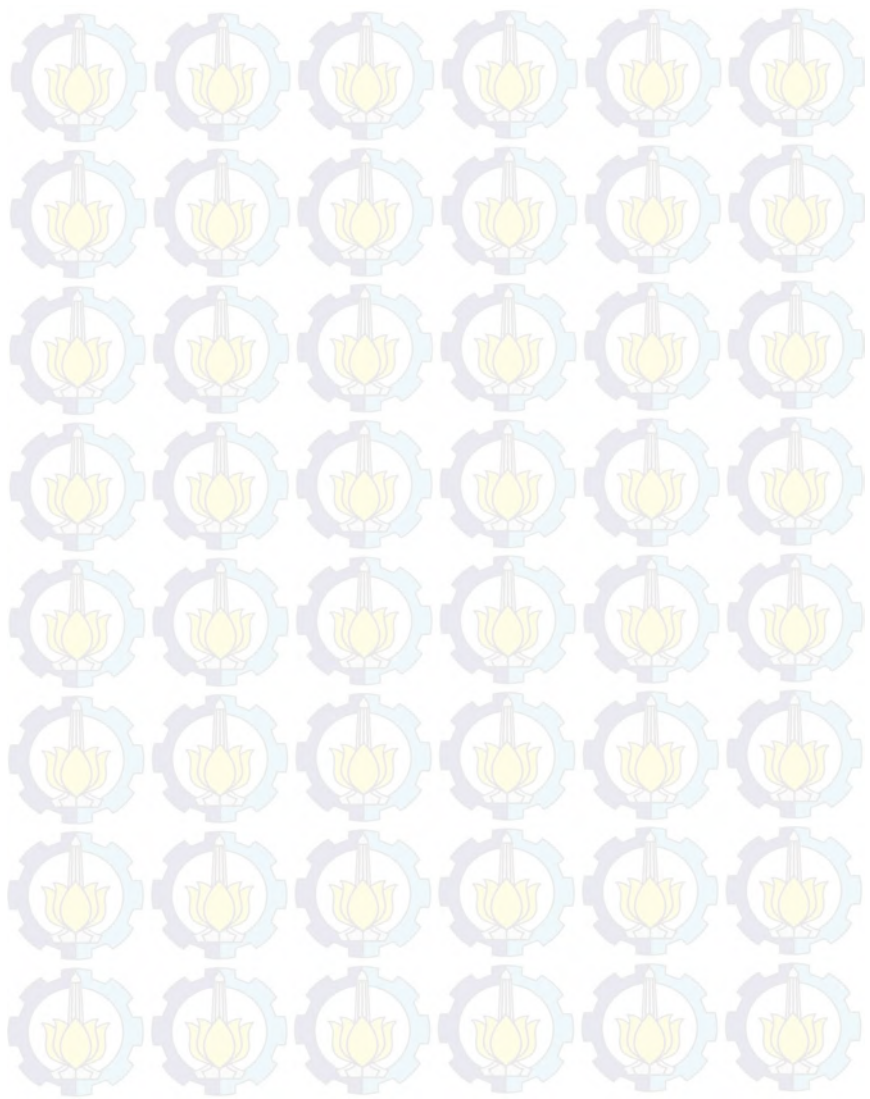


***Buku ini dipersembahkan untuk kedua orang tua tercinta
Untuk Bapak Sukardi yang tidak pernah lelah mendukung dan selalu
berusaha memberikan yang terbaik untuk penulis. Terimakasih Bapak atas
ketulusan dan kesabaranmu selama ini***

***Untuk Ibu Astiami, terimakasih selalu mendoakan dan selalu ada untuk
penulis, terimakasih atas kasih sayang dan doa yang tak pernah berhenti Ibu
panjatkan***

***Untuk kedua adik penulis, Dek Nabila dan Dek Zaki, terimakasih telah
memberikan hari – hari bahagia untuk kakak.***

Halaman ini sengaja dikosongkan.



KATA PENGANTAR

Puji syukur kepada Allah SWT karena dengan rahmat dan ridho-Nya penulis dapat menyelesaikan laporan tugas akhir yang berjudul “ANALISIS FAKTOR – FAKTOR YANG MEMPENGARUHI KESEMBUHAN DAN KEMATIAN PADA PASIEN STROKE DENGAN MENGGUNAKAN ASSOCIATION *RULE* MINING” sebagai salah satu syarat kelulusan di Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya.

Penulis berharap tugas akhir ini dapat memberikan manfaat bagi pembaca, menjadi salah satu batu loncatan bagi penulis untuk terus berkarya, dan memberikan sumbangsih bagi ilmu pengetahuan.

Penulis sadar bahwa dalam proses pengerjaan hingga terselesaikannya tugas akhir ini dibutuhkan bantuan dari berbagai pihak dan penulis mengucapkan terima kasih yang sebesar-besarnya kepada pihak-pihak yang membantu pengerjaan tugas akhir ini, antara lain:

- Seluruh Bapak dan Ibu dosen yang telah membagi banyak ilmu dan inspirasi kepada penulis.
- Ibu Wiwik Anggraeni, S.Si, M.Kom dan ibu Renny Pradina Kusumawardani, S.T., MT.yang telah dengan telaten dan sabar membimbing penulis dalam pengerjaan tugas akhir ini. Terimakasih atas waktu dan ilmu yang sudi dibagikan.
- Ibu Erma Suryani, ST, MT, Ph.D. dan Ibu Mahendrawati ER., ST., M.Sc., Ph.D. yang telah memberikan masukan dan saran untuk tugas akhir ini.
- Arif Surahman yang selalu ada dan memberikan semangat, terimakasih atas semuanya.
- Keluarga FOXIS yang selalu berjuang bersama-sama dan saling memberikan dukungan serta doa.
- Semua anggota lab SPK-IB dan admin lab tentunya, terimakasih atas dukungan dan do'anya selama ini.

Dan berbagai pihak yang tidak dapat penulis tuliskan namanya satu per satu. Semoga Tuhan membalas semua kebaikan yang telah dilakukan. Penulis menyadari banyaknya kekurangan yang dapat dikembangkan pada tugas akhir ini, oleh karena itu penulis menerima saran dan kritik yang membangun. Semoga tugas akhir ini dapat memberikan manfaat.

Surabaya, 10 Januari 2015

Penulis

DAFTAR ISI

DAFTAR ISI	xv
DAFTAR GAMBAR	xvii
DAFTAR TABEL	xix
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	3
1.3. Batasan Tugas Akhir	3
1.4. Tujuan Tugas Akhir	4
1.5. Manfaat Kegiatan Tugas Akhir	4
1.6. Sistematika Penulisan Tugas Akhir	4
BAB II TINJAUAN PUSTAKA	7
2.1. Penyakit Stroke	7
2.1.1. Jenis Penyakit Stroke	8
2.1.2. Penyebab	11
2.1.3. Teknik Diagnosis dan perawatan	11
2.2. <i>Association rule mining</i>	12
2.2.1. Frequent itemset generation dengan menggunakan algoritma apriori	15
2.2.2. Rule generation	16
2.3. Koefisien kontingensi	17
BAB III METODOLOGI Pengerjaan Tugas Akhir	19
3.1. Identifikasi Permasalahan	19
3.2. Pengumpulan Data	19
3.3. Studi Literatur	21
3.4. Proses <i>Association Rule Mining</i>	21
3.4.1. Data Preprocessing	21
3.4.2. Tahap Pelatihan	22
3.4.3. Tahap Pengujian	24
3.4.4. Evaluasi Performa dan Analisis hasil	24
3.5. Dokumentasi	25
BAB IV PENYIAPAN DATA	27
4.1. Profil data	27

4.2.	Tahapan Data Preprocessing	28
4.2.1.	Penambahan atau pengurangan atribut	33
4.2.2.	Pemilihan atribut yang relevan dengan koefisien kontingensi	37
4.2.3.	Gambaran hirarki data	40
4.2.4.	Konversi atribut bernilai kontinyu menjadi atribut diskrit	43
4.2.5.	Pembagian data <i>training</i> dan data <i>testing</i>	45
4.3.	Penyiapan file sesuai dengan format	48
BAB V TAHAP PELATIHAN DAN PENGUJIAN		49
5.1.	Penggalian <i>Rule</i>	49
5.2.	Pengujian <i>Rule</i>	53
BAB VI EVALUASI PERFORMA DAN ANALISIS HASIL		57
6.1.	Pengaruh perubahan minimum <i>support</i> dan minimum <i>confidence</i> pada kualitas <i>rule</i>	57
6.2.	Analisis <i>rule</i>	60
6.2.1.	Data Diagnosis	60
6.2.2.	Data Survival	74
6.2.3.	Pembobotan faktor – faktor yang berpengaruh terhadap data diagnosis dan data survival	89
BAB VII KESIMPULAN DAN SARAN		95
7.1.	Kesimpulan	95
7.2.	Saran	99
DAFTAR PUSTAKA		101
RIWAYAT PENULIS		105
LAMPIRAN		A-1

DAFTAR TABEL

Tabel 4-1 Detail data pasien stroke	28
Tabel 4-2 Tabel atribut data diagnosis yang dihilangkan.....	33
Tabel 4-3 Penggabungan atribut diagnosis.....	34
Tabel 4-4 Tabel atribut data survival yang dihilangkan.....	35
Tabel 4-5 Atribut data survival yang dihilangkan(2)	36
Tabel 4-6 Atribut terpilih untuk data diagnosis.....	37
Tabel 4-7 Atribut terpilih data survival.....	38
Tabel 4-8 Detail hirarki data	41
Tabel 5-1 Prosesntase kelas terhadap keseluruhan jumlah data dalam atribut.....	51
Tabel 6-1 <i>Rule</i> Stroke Ischaemic.....	60
Tabel 6-2 <i>Rule</i> stroke haemorrhagic.....	63
Tabel 6-3 <i>Rule</i> Indeterminate.....	67
Tabel 6-4 <i>Rule</i> Not a stroke.....	70
Tabel 6-5 <i>Rule</i> unknown stroke.....	73
Tabel 6-6 <i>Rule</i> dead.....	74
Tabel 6-7 <i>Rule</i> recover	77
Tabel 6-8 <i>Rule</i> dependent.....	81
Tabel 6-9 <i>Rule</i> not recover	83
Tabel 6-10 Faktor yang mempengaruhi diagnosis	89
Tabel 6-11 Faktor yang mempengaruhi kondisi akhir pasien	92

DAFTAR GAMBAR

Gambar 2-1 Penyebab kematian menurut data WHO	7
Gambar 2-2 Diagram alir association <i>rule</i> mining	16
Gambar 3-1 Metodologi pengerjaan tugas akhir	20
Gambar 3-2 Contoh pemangkasan <i>rule</i>	24
Gambar 4-1 Hirarki data secara umum	40
Gambar 4-2 AGE sebelum dikonversi	44
Gambar 4-3 Atribut AGE setelah dikonversi bernilai diskrit	44
Gambar 4-4 Diagram distribusi kelas atribut diagnosis	45
Gambar 4-5 Keterangan atribut diagnosis	46
Gambar 4-6 Atribut diagnosis pada data <i>training</i>	46
Gambar 4-7 Atribut diagnosis pada data <i>testing</i>	47
Gambar 4-8 OCCODE data <i>training</i>	47
Gambar 4-9 OCCODE data <i>testing</i>	48
Gambar 5-1 Potongan kode <i>load library</i>	49
Gambar 5-2 Potongan kode untuk <i>load data</i>	49
Gambar 5-3 Penerapan algoritma apriori	50
Gambar 5-4 proporsi data <i>training</i> diagnosis	51
Gambar 5-5 <i>Pruning rule</i>	52
Gambar 5-6 Menghitung <i>coverage</i>	53
Gambar 5-7 Potongan kode memasukkan <i>rule</i>	54
Gambar 5-8 Potongan kode <i>testing</i>	55
Gambar 6-1 Grafik pengukur performa <i>rule</i> diagnosis	58
Gambar 6-2 Grafik pengukur performa <i>rule</i> survival	59

BAB I PENDAHULUAN

Pada bab ini akan dijelaskan mengenai latar belakang, rumusan masalah, batasan pengerjaan, tujuan, dan manfaat dari pengerjaan tugas akhir.

1.1. Latar Belakang Masalah

Penyakit stroke merupakan salah satu penyakit yang memiliki resiko kematian tinggi, menurut data WHO penyakit ini merupakan penyakit penyebab kematian dengan jumlah terbesar ketiga bila dibandingkan dengan penyakit lain. Jumlah penderita stroke diseluruh dunia mencapai angka 15 juta orang, dari jumlah ini sebanyak 5 juta orang meninggal dan mengakibatkan kecacatan tubuh permanen bagi 5 juta orang lainnya (Mackay & Mensah). Di Indonesia jumlah penderita penyakit stroke meningkat dari tahun ke tahun, menurut riset yang dilakukan oleh Badan Penelitian dan Pengembangan Kesehatan Kementerian Kesehatan RI jumlah penderita meningkat dari 8,3 per 1000 penduduk (tahun 2007) menjadi 12,1 per 1000 penduduk (tahun 2013) (Badan Penelitian dan Pengembangan Kesehatan, 2013).

Pada penelitian sebelumnya banyak dilakukan analisis faktor - faktor yang dapat mempengaruhi seseorang dapat terjangkit stroke. Banyak faktor yang dapat dipertimbangkan seperti kondisi sebelum pasien terjangkit stroke, riwayat penyakit yang diderita pasien stroke dan berdasarkan jenis stroke yang diderita. Pada tugas akhir ini peneliti akan mencoba menganalisis faktor – faktor yang dapat mempengaruhi kesembuhan dan kematian pasien stroke pada saat pasien stroke dalam masa perawatan rumah sakit.

Pasien stroke yang menjadi objek penelitian sedang dalam masa perawatan rumah sakit. Faktor – faktor yang akan dianalisis berupa faktor – faktor mengenai kondisi dasar pasien sebelum perawatan rumah sakit (umur, jenis kelamin, kondisi fisik dan

gejala) serta tindakan medis yang dilakukan oleh rumah sakit terhadap pasien pada awal perawatan sampai pasien sembuh atau meninggal, masa pemantauan perawatan selama 14 hari dan *follow up* atau masa pemantauan pasien setelah keluar dari rumah sakit selama 6 bulan, data diperoleh dari *database The International Stroke Trial*.

Metode yang digunakan dalam penelitian ini adalah *association rule mining*, dimana metode ini dapat menghasilkan keluaran berupa pola yang dapat menggambarkan hubungan atau keterkaitan antar banyak faktor. Pola tersebut tidak bisa didapatkan melalui metode lain seperti korelasi dan regresi. Misalnya saja dengan menggunakan korelasi kita hanya dapat mengetahui bahwa pasien yang memiliki tekanan darah tinggi memiliki peluang tinggi untuk meninggal, namun kita tidak dapat mengetahui apakah jika faktor tekanan darah di gabungkan dengan faktor lain seperti usia pasien akan tetap menghasilkan keluaran yang sama.

Metode *association rule mining* sesuai untuk diterapkan karena banyaknya atribut atau faktor yang terdapat di dalam data. Metode ini dapat membantu memenuhi tujuan penelitian, yaitu untuk menggali pola – pola yang ada di dalam data sehingga dapat dianalisis faktor – faktor yang memiliki hubungan atau berkontribusi dalam kesembuhan dan kematian pasien stroke. Pada penelitian sebelumnya, *association rule mining* digunakan untuk mengetahui faktor apa saja yang dapat menimbulkan resiko seseorang terjangkit penyakit jantung (pada laki – laki dan perempuan), peneliti tersebut mencoba untuk menggunakan tiga algoritma sekaligus, yaitu apriori, *apriori prediction* dan tertius, hasilnya ketiga algoritma menghasilkan kesimpulan yang sama (Nahar, Imam, Tickle, & Chen, 2013). Pada penelitian tugas akhir ini dipilih algoritma apriori karena dibandingkan dengan algoritma lain pada penelitian sebelumnya, algoritma apriori dapat melakukan pencarian dalam waktu yang lebih singkat mengingat data yang akan digunakan cukup besar (Nahar, Imam, Tickle, & Chen, 2013). Hasil dari tugas akhir ini diharapkan dapat membantu

penanganan pasien stroke dengan lebih tepat sehingga dapat mengurangi kematian pasien rumah sakit akibat penyakit stroke.

1.2. Rumusan Masalah

Dari uraian pada bagian latar belakang, maka rumusan permasalahan pada pengerjaan tugas akhir ini adalah faktor – faktor apa saja yang mempengaruhi kesembuhan dan kematian pada pasien stroke, sehingga hasilnya dapat digunakan sebagai langkah awal untuk penanganan pasien stroke yang lebih baik. Untuk mencari faktor – faktor tersebut ada tahapan yang harus dilakukan, dan dibawah ini adalah sub permasalahan yang harus diselesaikan :

- a. Bagaimana cara penyiapan data sehingga data dapat dipakai dalam proses *association rule mining*?
- b. Berapakah nilai minimum *support* dan nilai minimum *confidence* yang akan dipakai dalam proses *association rule mining*?

1.3. Batasan Tugas Akhir

Pengerjaan tugas akhir ini memiliki beberapa batasan antara lain sebagai berikut:

- a. Tugas akhir ini menggunakan studi kasus pasien stroke yang dirawat di rumah sakit, data terdiri dari 19.435 pasien yang berasal dari berbagai rumah sakit yang berasal dari 43 negara (data mengenai detail identitas pasien dan rumah sakit yang terlibat dirahasiakan).
- b. Data diambil dari database *International Stroke Trial*. Data pasien yang akan digunakan merupakan data pasien stroke mulai dari tahun 1991 sampai dengan tahun 1996.
- c. Data yang digunakan berisi data pasien yang dipantau pada masa perawatannya dan mendapatkan pengawasan selama 6 bulan setelah perawatan di rumah sakit.

1.4. Tujuan Tugas Akhir

Tujuan dari pengerjaan tugas akhir ini adalah mengetahui faktor – faktor yang mempengaruhi kesembuhan dan kematian pasien stroke sehingga dapat digunakan untuk membantu penanganan pasien stroke agar lebih baik, dalam arti mengurangi peluang kematian dan meningkatkan peluang kesembuhan dengan mempertimbangkan faktor – faktor yang berpengaruh terhadap keduanya .

1.5. Manfaat Kegiatan Tugas Akhir

Manfaat yang diharapkan akan didapatkan setelah pengerjaan tugas akhir ini adalah sebagai berikut:

- a. Hasil dari tugas akhir ini diharapkan dapat membantu bidang kedokteran dalam proses penanganan pasien stroke yang tepat.
- b. Sebagai referensi untuk mengetahui bagaimana cara menerapkan *association rule mining* dan algoritma apriori.
- c. Sebagai masyarakat dapat mengetahui informasi mengenai faktor – faktor yang berpengaruh pada kesembuhan dan kematian pasien stroke, sehingga dapat melakukan tindakan preventif dan tepat sesuai dengan kapasitas yang dimiliki.

1.6. Sistematika Penulisan Tugas Akhir

Sistematika penulisan Laporan Tugas Akhir ini dibagi menjadi tujuh bab sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang, rumusan masalah, batasan pengerjaan tugas akhir, manfaat tugas akhir, tujuan tugas akhir dan sistematika penulisan yang diterapkan dalam memaparkan tugas akhir.

BAB II TINJAUAN PUSTAKA

Pada bab ini dijelaskan mengenai teori-teori terkait yang bersumber dari buku, jurnal, ataupun artikel yang berfungsi sebagai dasar dalam melakukan pengerjaan tugas akhir agar dapat memahami konsep atau teori penyelesaian permasalahan yang ada. Pada bab ini terdapat uraian mengenai penyakit stroke, uraian mengenai *association rule mining* dan algoritma apriori.

BAB III METODE Pengerjaan Tugas Akhir

Pada bab ini dijelaskan mengenai langkah–langkah pengerjaan tugas akhir beserta metode yang digunakan. Langkah–langkah pengerjaan dijelaskan dalam sebuah diagram alur yang sistematis dan akan dijelaskan tahap demi tahap.

BAB IV PENYIAPAN DATA

Pada bab ini akan dijelaskan bagaimana tahapan awal dari proses *association rule mining* yaitu *data preprocessing*.

BABV TAHAP PELATIHAN DAN PENGUJIAN

Pada bab ini akan dijelaskan mengenai penggalian *rule* dari data *training* atau data latih dan bagaimana *rule* tersebut diimplementasikan di dalam data *testing* atau data uji sehingga bisa diketahui akurasi *rule* yang ada

BAB V EVALUASI PERFORMA DAN ANALISIS HASIL

Pada bab ini akan dilakukan analisis kualitas *rule* yang telah diuji ke data *testing* dan pengaruh minimum *support* dan *confidence* terhadap kualitas *rule*. Selain itu akan dianalisis pola yang terbentuk di dalam *rule* dan diterjemahkan sehingga dapat diketahui faktor – faktor yang berpengaruh terhadap kesembuhan dan kematian pasien stroke.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari pengerjaan tugas akhir dan saran untuk pengembangan lebih lanjut.

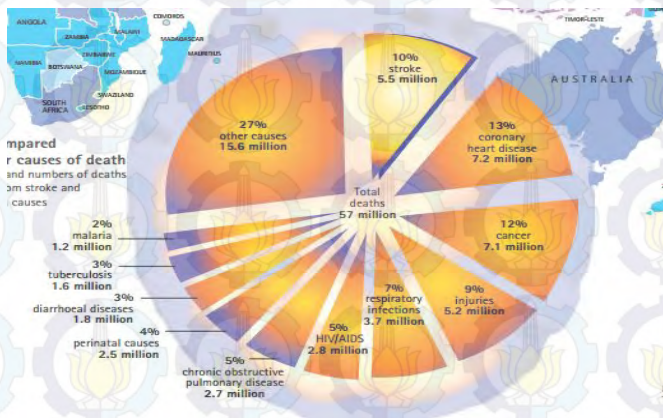
BAB II

TINJAUAN PUSTAKA

Pada bagian bab ini akan dijelaskan mengenai berbagai teori yang digunakan sebagai dasar dan acuan dalam pengerjaan tugas akhir ini.

2.1. Penyakit Stroke

Penyakit stroke disebabkan oleh terganggunya aliran darah atau asupan darah ke otak, gangguan tersebut disebabkan oleh pembuluh darah yang pecah atau tersumbat. Hal ini dapat mengakibatkan hilangnya asupan oksigen dan nutrisi – nutrisi yang dibutuhkan otak sehingga otak dapat mengalami kerusakan. Efek dari penyakit ini berbeda – beda tergantung dari bagian otak mana yang mengalami kerusakan, serangan stroke yang parah dapat mengakibatkan kematian mendadak. Dibandingkan dengan penyakit lain, penyakit stroke memiliki resiko kematian yang lebih tinggi menurut referensi WHO presentase kematian akibat penyakit ini dan perbandingannya dengan penyakit lain dapat dilihat pada gambar 2-1.



Gambar 2-1 Penyebab kematian menurut data WHO

2.1.1. Jenis Penyakit Stroke

2.1.1.1. Stroke ringan (TIA)

Stroke ringan atau transient ischemic attack (TIA) terjadi ketika asupan oksigen ke otak terhalang sebentar, lalu kembali normal jika asupan terhalang lebih dari 24 jam, maka akan dikategorikan sebagai stroke biasa.

2.1.1.2. Stroke Iskemik

Stroke digolongkan menjadi stroke iskemik jika pembuluh darah yang memasok darah ke otak tersumbat, stroke jenis ini yang umumnya diderita oleh pasien (hampir 90% stroke yang ada adalah iskemik). Penyebab stroke iskemik adalah penumpukan lemak yang melapisi dinding pembuluh darah sampai lapisan tersebut menjadi plak, plak mengakibatkan darah sulit mengalir dan dapat menyebabkan trombus atau bekuan darah.

Stroke iskemik dibedakan menjadi dua, yang pertama adalah stroke trombotik yang disebabkan oleh trombus yang ada di pembuluh darah otak yang sudah sempit. Yang kedua adalah stroke embolik, disebabkan oleh trombus, gelembung udara atau pecahan lemak (emboli) yang terbentuk di bagian tubuh lain, misalnya jantung, bagian dada dan leher lalu terbawa aliran darah ke otak. Kelainan jantung yang disebut fibrilasi atrium dapat menyebabkan trombus mengalir ke otak.

2.1.1.3. Stroke Hemoragik

Stroke hemoragik penyebabnya adalah pembuluh darah yang bocor atau pecah di dalam atau di sekitar otak sehingga asupan darah ke bagian otak yang dituju menjadi terhenti. Karena

pembuluh darah yang pecah atau bocor, darah akan membanjiri dan dapat memampatkan jaringan otak disekitar daerah kebocoran, akibatnya fungsi otak terganggu bahkan mati.

Dua jenis stroke hemoragik, dr Dewi menuliskan bahwa dua jenis stroke ini dibedakan berdasarkan lokasi pendarahan. Yang pertama disebabkan oleh perdarahan intraserebral, perdarahan intraserebral adalah perdarahan yang terjadi di dalam otak, disebabkan oleh cedera otak atau kelainan pembuluh darah (aneurisma atau angioma), pendarahan ini juga bisa disebabkan oleh darah tinggi yang kronis. Yang kedua disebabkan oleh perdarahan subarachnoid yaitu perdarahan dalam ruang subarachnoid, ruang di antara lapisan dalam (Pia mater) dan lapisan tengah (arachnoid mater) dari jaringan selaput otak (meninges). Penyebab paling umum adalah pecahnya tonjolan (aneurisma) dalam arteri. Perdarahan subarachnoid dapat menyebabkan cacat permanen atau kematian.

2.1.1.4. Subtipe penyakit stroke

Di dalam data penelitian, penyakit stroke juga digolongkan berdasarkan gejala yang diderita pasien, diantaranya adalah:

a. *Total Anterior Circulation Stroke (TACS)*

Gejala yang diderita pasien antara lain:

- Gangguan kemampuan berbahasa, contohnya pasien mengalami kesulitan untuk memahami arti kata.
- Gangguan berkomunikasi / berbicara.

- Pasien sulit untuk mengenali ruang , tempat atau lingkungan.
- Tingkat kesadaran pasien menurun.
- Terjadi pelemahan pada sebagian tubuh, biasanya terjadi pada satu sisi tubuh.

b. *Partial Anterior Circulation Stroke (PACS)*

Pasien dapat digolongkan menderita penyakit stroke dengan subtype ini apabila mengalami paling tidak 2/3 dari gejala stroke TACS.

c. *Lacunar Stroke (LACS)*

Pada stroke lacunar gejala yang diderita pasien biasanya berupa gangguan kemampuan motorik dan sensorik. Pasien juga merasakan adanya pelemahan pada bagian lengan atau kaki, pada kebanyakan kasus pelemahan lebih sering menyerang bagian kaki dari pada bagian lengan.

d. *Posterior Circulation (POCS)*

Secara klinis pasien mengalami gejala kelumpuhan saraf motorik dan sensorik. Terdapat cacat atau masalah pergerakan mata, sehingga pasien merasa kehilangan sebagian visual dari objek yang dilihat. Hal tersebut dapat terjadi pada salah satu atau kedua mata. Gejala lain yang dialami pasien adalah gangguan pendengaran pada salah satu atau kedua telinga.

2.1.2. Penyebab

Banyak faktor – faktor yang menyebabkan seseorang menderita penyakit stroke (faktor resiko). Rambe (2006) membagi faktor tersebut menjadi dua yaitu faktor genetik dan gaya hidup.

- Faktor genetik (tidak bisa dimodifikasi) : Usia, ras, jenis kelamin, riwayat penyakit stroke dalam keluarga serta sejarah penyakit stroke yang pernah diderita.
- Gaya hidup : Riwayat penyakit darah tinggi, diabetes mellitus, merokok, hyperlipidemia serta konsumsi makanan dan minuman (bahan pengawet, alkohol, sintetis, dan lain – lain).

2.1.3. Teknik Diagnosis dan perawatan

Dalam penelitian tugas akhir ini pasien yang terkena stroke atau menunjukkan gejala stroke diperiksa secara menyeluruh, seperti melakukan CT scan, pemberian obat aspirin dan heparin, pengecekan fungsi – fungsi tubuh mana yang terganggu, contohnya tangan tidak dapat digerakkan pada satu sisi tubuh, gangguan penglihatan atau bahkan pasien sudah dalam keadaan tidak sadar atau pingsan.

Setelah itu pengumpul data mulai merekam data pasien yang berupa data umum pasien (jenis kelamin dan umur) dan data kondisi pasien sebelum perawatan dan hasil diagnosis, setelah itu pasien akan dipantau selama 14 hari pertama, dalam jangka waktu tersebut apakah ada kejadian seperti komplikasi, pembedahan atau operasi, kematian, dan lain – lain. Setelah 14 hari akan direkam apakah pasien sudah dapat meninggalkan rumah sakit atau harus dipindah ke unit lain beserta kondisinya, masih ada pemantauan lanjut selama 6 bulan dan selama kurun waktu tersebut akan direkam hasil akhir kondisi pasien. Namun ada beberapa kasus dengan data pasien yang tidak lengkap

karena pasien meninggal atau melokak salah satu proses perawatan yang diberikan rumah sakit. (Sandercock, Niewada, & Czlonkowska, 2011)

2.2. *Association rule mining*

Association rule mining merupakan salah satu dari prosedur data mining yang banyak digunakan. Berfungsi untuk mencari pola yang sering muncul diantara banyak transaksi atau untuk menemukan *rule* atau aturan antara suatu kombinasi item.

Rule berbentuk $X \rightarrow Y$, dimana X dan Y dapat berisi lebih dari satu item. X dan Y merupakan *disjoint itemset* ($X \cap Y = \emptyset$). *Rule* tersebut menggambarkan bahwa Y akan terjadi ketika X terjadi, contohnya seseorang yang membeli buku akan membeli pensil dan juga bolpoin, maka penulisan *rule*nya adalah sebagai berikut:

$$\{\text{buku}\} \rightarrow \{\text{pensil, bolpoin}\}$$

Association rule mining akan menghasilkan *rule* yang menentukan seberapa besar hubungan atau pengaruh antara X dan Y. Untuk mengukur kualitas output akan digunakan beberapa pengukur, diantaranya adalah sebagai berikut:

- *Support*

Support adalah presentase dari transaksi yang mengandung itemset (X dan Y).

$$\text{Support}(X \rightarrow Y) = \frac{\text{Jumlah transaksi yang mengandung } X \text{ dan } Y}{\text{Jumlah transaksi keseluruhan}} \dots (2.1)$$

Minimum support menggambarkan batas ambang prosentase kombinasi item dalam transaksi database.

- **Confidence**

Merupakan rasio antara jumlah transaksi yang berisi X dan Y dan jumlah transaksi yang berisi X. Minimum *confidence* menggambarkan batas ambang kuatnya hubungan antar item dalam aturan asosiatif.

$$Confidence(X \rightarrow Y) = \frac{\text{Jumlah transaksi yang mengandung X dan Y}}{\text{Jumlah transaksi yang mengandung X}} \dots (2.2)$$

Minimum confidence merupakan batas ambang nilai *confidence* yang harus dipenuhi .

- **Lift**

Lift mengukur tingkat kepentingan dari suatu *rule*.

$$\text{lift}(X \rightarrow Y) = \text{lift}(Y \rightarrow X) = \frac{\text{conf}(X \rightarrow Y)}{\text{supp}(Y)} = \frac{\text{conf}(Y \rightarrow X)}{\text{supp}(X)} \dots (2.3)$$

- **Coverage**

Mengukur seberapa besar cakupan implementasi *rule* pada keseluruhan data. Dapat dihitung dari nilai *support* lhs (*left hand side*).

Istilah lain yang berhubungan dengan **association rule mining** yang banyak dipakai penulis dalam tugas akhir ini diantaranya adalah sebagai berikut :

• **Atribut**

Berisi nilai yang menjelaskan karakteristik dari obyek yang ada di dalam data (dalam tugas akhir ini obyek data adalah pasien). Dalam *association rule mining*, nilai yang terkandung dalam suatu atribut dapat disebut dengan item.

- **Kelas**

Merupakan label yang ada di dalam suatu atribut atau biasa disebut dengan kategori.

- **Itemset**

Merupakan kumpulan dari item, jumlah item di dalam itemset minimal satu.

- **Transaksi**

Merupakan nama lain dari *tuple* atau *record*, dalam tugas akhir ini satu *record* mewakili satu pasien, jadi setiap transaksi merupakan transaksi yang dilakukan oleh pasien yang berbeda – beda.

- **Rule pruning,**

Merupakan proses pemangkasan *rule* yang bertujuan untuk menghilangkan *rule* yang duplikat atau redundan.

Pada penelitian sebelumnya Association *rule* mining digunakan untuk mengetahui faktor apa saja yang dapat menimbulkan resiko seseorang terjangkit penyakit jantung (pada laki – laki dan perempuan), peneliti tersebut mencoba untuk menggunakan tiga algoritma sekaligus, yaitu apriori, *apriori prediction* dan tertius. Hasil dari ketiga algoritma tersebut relatif sama sehingga kesimpulan yang ditarik dari hasil keduanya pun sama (Nahar, Imam, Tickle, & Chen, 2013), namun pada prosesnya algoritma yang paling cepat mengeluarkan hasil adalah algoritma apriori, oleh karena itu pada tugas akhir ini dipilih algoritma apriori dalam proses *Frequent Itemset Generation*.

2.2.1. Frequent itemset generation dengan menggunakan algoritma apriori.

Frequent itemset adalah itemset yang memiliki *support* yang lebih besar atau sama dengan batas ambang minsup (minimum *support*). (Tan, Steinbach, & Kumar, 2005)

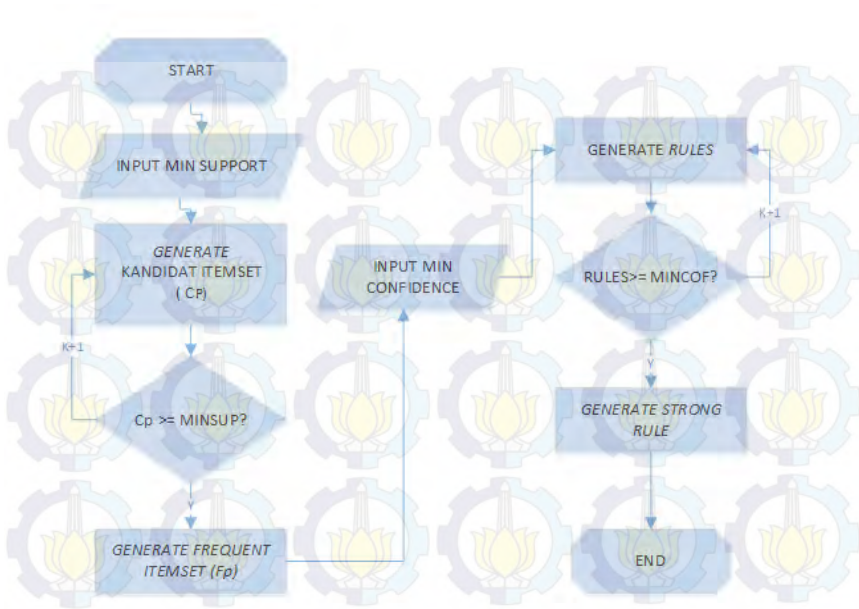
Algoritma Apriori

Pada tugas akhir ini akan diterapkan algoritma apriori karena dari penelitian sebelumnya algoritma ini membutuhkan waktu lebih sedikit untuk menyelesaikan prosesnya dibandingkan dengan algoritma yang lain. Sebelum menjalankan algoritma ini, parameter minsup atau persentase kemunculan suatu *rules* harus ditentukan terlebih dahulu. Selain itu, minimum *confidence* atau persentase kemunculan *rules* terhadap kemunculan anteseden-nya juga harus ditentukan terlebih dahulu.

Algoritmanya adalah sebagai berikut :

1. misal $k=1$
2. bentuk frequent item sets yang terdiri dari k -item
3. ulangi hingga tidak ada lagi frequent item sets yang baru
4. bentuk kandidat item sets dengan panjang $k+1$ dari frequent item sets dengan panjang k
5. buang kandidat item sets yang berisi subset dengan panjang k yang tidak frequent
6. hitung *support* dari setiap kandidat dengan scanning basisdata
7. eliminasi kandidat yang infrequent

Gambar 2-2 menjelaskan diagram alir dari proses association *rule* mining. Dimana setelah menemukan frequent itemset akan dilanjutkan ke proses *rule* generation.



Gambar 2-2 Diagram alir association rule mining

2.2.2. Rule generation

Setelah mendapatkan frequent itemset hasil dari tahapan sebelumnya, tahap selanjutnya adalah mendapatkan *rule* yang memenuhi *confidence* (lebih besar atau sama dengan batas ambang yang ditentukan). Tidak perlu dilakukan penghitungan *support* karena *rules* yang akan diproses dalam tahap ini telah memenuhi *minsup* yang telah ditentukan. (Tan, Steinbach, & Kumar, 2005).

2.3. Koefisien kontingensi

Uji koefisien kontingensi berfungsi untuk mengujur keeratan hubungan antara dua atribut. Uji ini diperlukan dalam tahap *data preprocessing*. Dengan melakukan uji koefisien kontingensi diharapkan semua atribut yang diuji merupakan atribut yang relevan.

Koefien kontingensi dipilih karena data yang ada memiliki distribusi yang tidak normal, atribut data berisi data kategorik dan dalam satu atribut ada yang memiliki lebih dari dua kategori. Berikut ini adalah formula dari koefisien kontingensi:

$$c = \sqrt{\frac{\chi^2}{N + \chi^2}} \quad \dots(2.4)$$

Dimana :

$$\chi^2 = \sum_{i=1}^b \sum_{j=1}^k \frac{D_{ij}^2}{E_{ij}}, \text{ derajat bebas (df) } = (b-1)*(k-1)$$

$$D_{ij} = O_{ij} - E_{ij} \quad \dots(2.5)$$

- N = total banyaknya observasi
- c = koefisien kontingensi
- b = banyaknya baris pada tabel kontingensi (crosstab)
- k = banyaknya kolom pada tabel kontingensi(crosstab)
- i = 1, 2, ..., b
- j = 1, 2, ..., k
- O_{ij} = data observasi baris ke-i kolom ke-j pada tabel kontingensi
- E_{ij} = nilai frekuensi harapan ke-ij untuk O_{ij}
- χ^2 = Chi-square hasil perhitungan

BAB III

METODOLOGI Pengerjaan Tugas Akhir

Tugas akhir ini diselesaikan melalui berbagai tahap pengerjaan yang dapat dilihat pada gambar 3-1. Pada gambar tersebut juga dijelaskan keluaran dari masing-masing tahap yang terlihat pada bagian kanan dari alur pengerjaan.

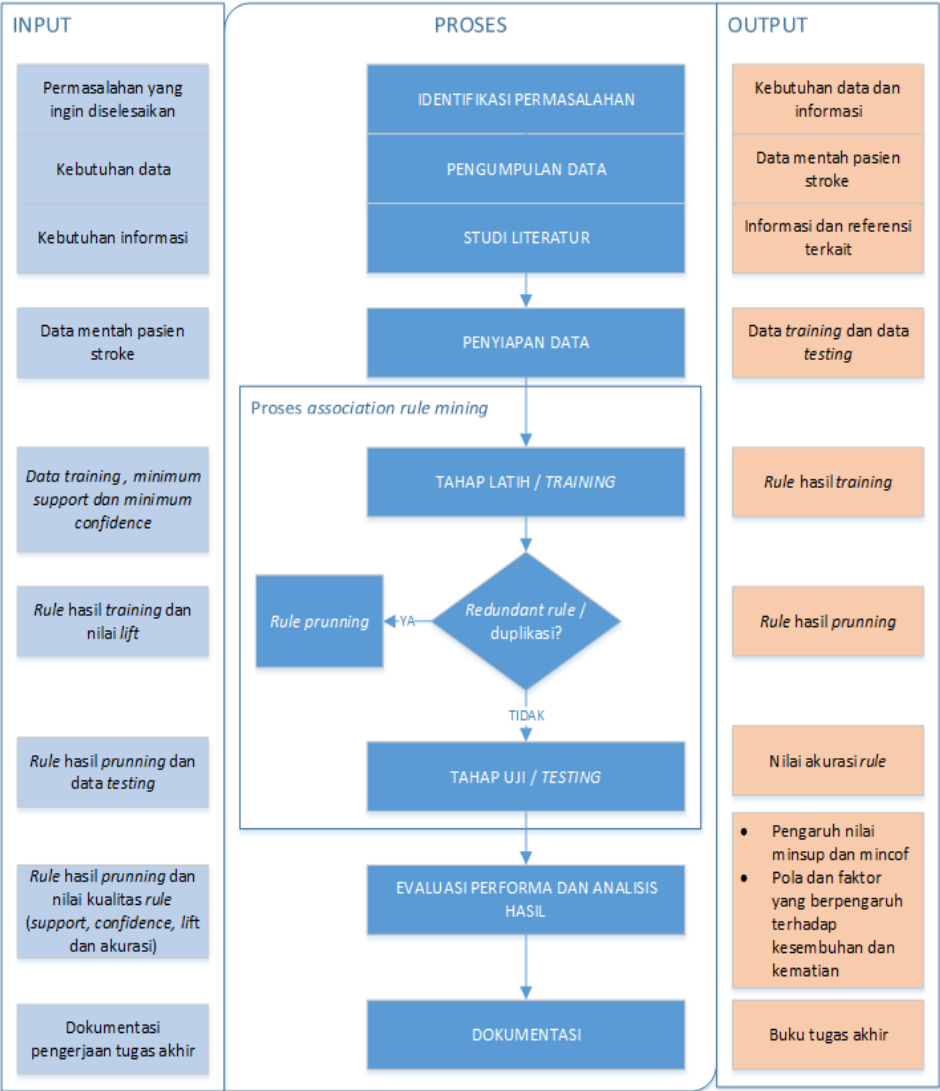
Berikut ini merupakan penjelasan dari tahapan pengerjaan tugas akhir yang telah digambarkan pada gambar 3-1.

3.1. Identifikasi Permasalahan

Tahapan ini bertujuan untuk mengetahui permasalahan yang akan diselesaikan dalam penelitian tugas akhir. Dalam penelitian sebelumnya banyak peneliti yang menganalisis faktor – faktor yang mempengaruhi resiko seseorang untuk terjangkit stroke, namun belum ada penelitian yang membahas mengenai faktor – faktor yang mempengaruhi kesembuhan dan kematian seseorang setelah terjangkit penyakit stroke, oleh karena itu permasalahan tersebut akan dibahas pada penelitian tugas akhir ini .

3.2. Pengumpulan Data

Dalam tahapan ini akan dikumpulkan data dan informasi yang dibutuhkan dalam pengerjaan tugas akhir, yaitu data pasien rawat inap rumah sakit yang menderita stroke. Pengumpulan informasi dilakukan dengan cara *browsing* . Data yang didapatkan adalah data pasien stroke, pasien yang terekam sebanyak 19.435 orang, berasal dari rumah sakit yang tersebar di 43 negara. Data pribadi pasien seperti nama dan alamat pasien dirahasiakan begitu juga dengan data rumah sakit yang terlibat. Data diambil dari database International Stroke Trial, dan telah diizinkan untuk digunakan secara publik.



Gambar 3-1 Metodologi pengerjaan tugas akhir

Data berisi detail mengenai kondisi pasien pada saat sebelum perawatan dilakukan, pada saat perawatan dan pada saat pemantauan sesudah perawatan. Pemantauan masa perawatan dilakukan selama 14 hari, dan pemantauan sesudah perawatan dilakukan selama 6 bulan. Pada saat sebelum perawatan dimulai data yang dikumpulkan adalah data mengenai kondisi pasien, seperti gejala yang di derita, informasi dasar pasien (umur, jenis kelamin) dan data mengenai hasil tindakan sebelum diagnosis. Data yang direkam selama perawatan adalah data mengenai tindakan medis yang dilakukan terhadap pasien (misalnya pemberian obat atau pembedahan) dan kejadian apa saja yang terjadi selama masa 14 hari (sudah bisa dipulangkan dari rumah sakit, meninggal, komplikasi, kambuh dan lain – lain) serta kondisi tubuh pasien . Data setelah perawatan merupakan hasil pantauan selama 6 bulan terhadap pasien, hasil yang direkam adalah kesembuhan (sembuh total atau sembuh tetapi mengalami cacat tubuh), belum sembuh dan kematian..

3.3. Studi Literatur

Pada tahapan ini akan dilakukan penggalian teori – teori yang relevan dengan permasalahan dan kebutuhan pengerjaan tugas akhir. Studi literature ini meliputi pemahaman mengenai *Assosiation rule mining* dan algoritma apriori serta bagaimana penerapannya untuk menyelesaikan permasalahan tugas akhir.

3.4. Proses Association Rule Mining

Setelah data dan referensi terkumpul, akan dilanjutkan dengan proses *association rule mining*. Dalam proses ini ada beberapa tahapan yang harus dilakukan, tahapan tersebut tergambar dalam flowhart dibawah ini :

3.4.1. Data Preprocessing

Pada tahapan ini, data akan dibagi menjadi 2 kelompok, yaitu:

- a. Kelompok pertama merupakan data awal pasien sebelum perawatan (data diagnosis), berisi kondisi awal pasien seperti usia, jenis kelamin dan gejala yang diderita sampai dengan atribut diagnosis. Atribut diagnosis berisi jenis stroke yang diderita pasien.
- b. Kelompok kedua merupakan data yang dikumpulkan selama 14 hari perawatan sampai 6 bulan masa pemantauan (data survival), isi data berupa kondisi pasien selama perawatan, tindakan medis yang diberikan kepada pasien selama perawatan sampai dengan kondisi terakhir pasien setelah 6 bulan perawatan.

Detail mengenai atribut dan isi atribut akan dijelaskan pada bab selanjutnya. Proses lain dalam tahapan ini adalah pemilihan atribut berdasarkan koefisien kontingensi, pembagian data *training* dan data *testing*, penambahan atau pengurangan atribut, serta konversi data *continuous* menjadi diskrit atau kategorikal.

3.4.2. Tahap Pelatihan

Pada tahapan ini akan digunakan data *training*, ketiga kelompok data akan melalui tahapan yang sama. Yaitu :

a. Mencari *Frequent Itemset* dengan menggunakan algoritma apriori

Menghasilkan *frequent itemset* yang memenuhi batas ambang minsup. Penentuan nilai minimum *support* dan minimum *confidence* sebaiknya tidak terlampau kecil, besar dan kecil nilai minimum *support* dan *confidence* sangat bergantung pada ukuran database yang diolah (Budi, 2010). Oleh karena itu, minimum *support* dan *confidence* akan ditentukan setelah *data preprocessing* selesai, sehingga kita bisa mengetahui secara pasti besarnya data yang akan diproses ke tahapan ini.

Pada tugas akhir ini akan digunakan nilai minimum *support* dan minimum *confidence* dengan interval tertentu (rentang awal yang besar dan kemudian akan dipersempit sesuai dengan kebutuhan). Dengan eksperimen ini diharapkan peneliti dapat memperoleh nilai *minsup* dan *minconf* yang optimal. Cara ini pernah dilakukan pada penelitian sebelumnya yaitu mengenai penggunaan algoritma apriori dalam *content based recommender system* (Oktoria, Maharani, & Firdaus, 2010). Penentuan minimum *support* dan *confidence* akan mempengaruhi jumlah *rule* yang dihasilkan. Semakin besar nilai dari keduanya, *rule* yang dikeluarkan akan semakin sedikit. Pada penelitian ini diharapkan *rule* yang dihasilkan memiliki jumlah yang seoptimal mungkin. Pada proses analisis hasil akan dilakukan validasi, dalam proses tersebut kita dapat melihat minimum *support* dan *confidence* berapakah yang dapat menghasilkan akurasi yang tinggi.

b. Rule Generation

Menghasilkan *rule* yang memenuhi batas ambang *minimum confidence*. Detail dari proses mencari *frequent itemset* dan *rule generation* ini telah dijelaskan dalam bagian sebelumnya, yaitu bagian tinjauan pustaka.

Setelah *rule* terbentuk, akan disaring kembali untuk menghindari duplikasi *rule*. Tahap ini disebut dengan *pruning* atau pemangkasan. Pemangkasan akan dilakukan terhadap *rule* yang tidak memberikan tambahan pengetahuan. Contohnya adalah seperti *rule* pada gambar 3-2. *Rule* 1 menyatakan bahwa semua anak – anak yang berada di kelas dua statusnya selamat, pada saat kita melihat *rule* ke 2, tidak ada tambahan pengetahuan lagi karena pada *rule* sebelumnya sudah dinyatakan bahwa semua anak – anak di kelas dua selamat. Jika *rule* 2 merupakan *super rule* dari *rule* 1 dan nilai *lift rule* 2 lebih kecil atau sama dengan nilai *lift rule* 1 maka *rule* 2 dinyatakan *redundant*. Oleh karena itu *rule* 2 harus dipangkas. (Association rules, 2014)

	lhs	rhs	support	confidence	lift
1	{Class=2nd, Age=Child}	=> {Survived=Yes}	0.010904134	1.0000000	3.095640
2	{Class=2nd, Sex=Female, Age=Child}	=> {Survived=Yes}	0.005906406	1.0000000	3.095640

Gambar 3-2 Contoh pemangkasan *rule*

3.4.3. Tahap Pengujian

Informasi yang dibutuhkan dalam penelitian ini adalah informasi mengenai faktor – faktor apa saja yang berhubungan atau dapat mempengaruhi kesembuhan dan kematian penderita stroke, faktor – faktor tersebut dapat diketahui dari keluaran *rule generation* pada tahapan sebelumnya. Untuk itu kita akan melakukan evaluasi dengan mengukur kualitas *rule* yang sudah terbentuk, akan dilakukan penggalan pada data *testing*.

Setelah *rule* yang terbentuk dari data *training* di terapkan di data *testing*, kita dapat menghitung nilai akurasi dan *coverage* dari masing – masing *rule* terhadap data *testing*.

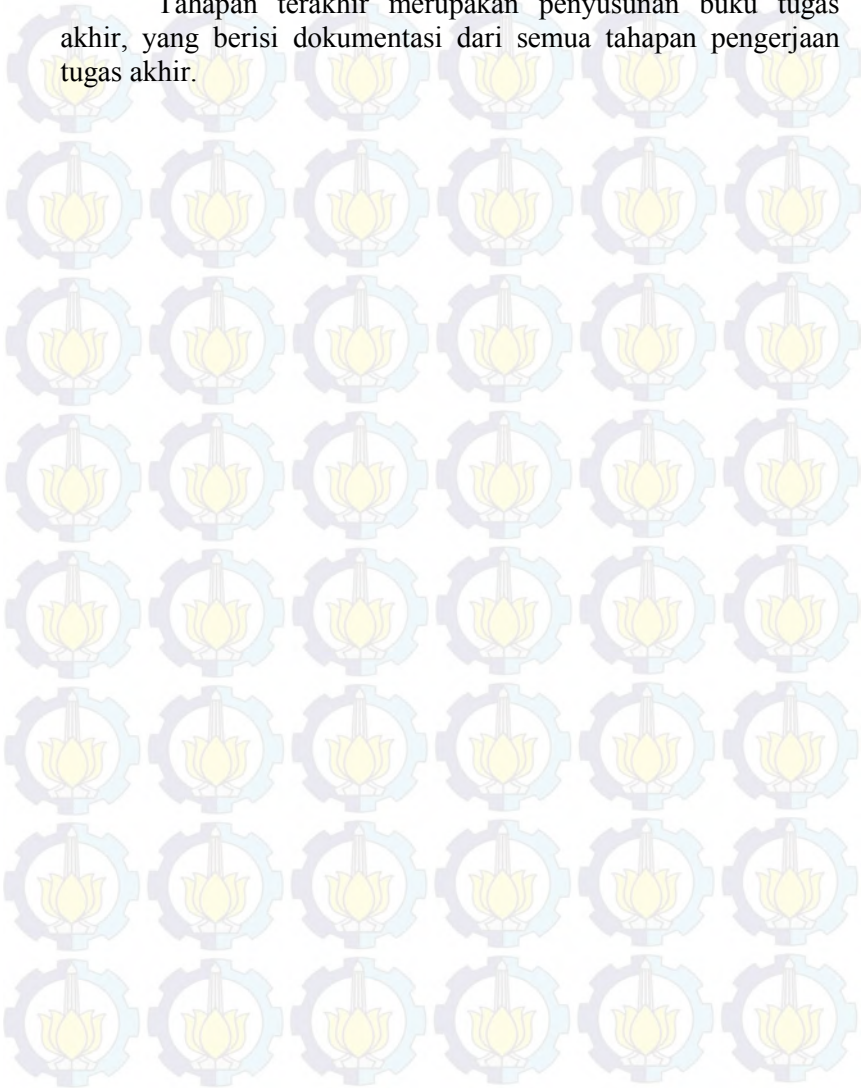
3.4.4. Evaluasi Performa dan Analisis hasil

Dalam tahapan ini dilakukan analisis hasil yang bertujuan untuk memperoleh relevansi informasi dari permasalahan yang ada di dalam tugas akhir. Ada dua analisis yang akan dilakukan yaitu :

- Analisis pengaruh perubahan minsup dan mincof terhadap kualitas *rule* yang dihasilkan.
- Rule* akan dianalisis dan diterjemahkan sehingga kita dapat mengetahui faktor – faktor apa saja yang berpengaruh terhadap kesembuhan dan kematian pasien stroke.

3.5. Dokumentasi

Tahapan terakhir merupakan penyusunan buku tugas akhir, yang berisi dokumentasi dari semua tahapan pengerjaan tugas akhir.



BAB IV

PENYIAPAN DATA

Pada bab ini akan dijelaskan mengenai proses penyiapan data, yang diawali dengan penjelasan mengenai profil data dan akan dilanjutkan dengan tahapan – tahapan penyiapan data (*data preprocessing*).

4.1. Profil data

Data berisi data pasien stroke yang menjalani rawat inap di rumah sakit.. Data yang didapatkan adalah data pasien stroke, pasien yang terekam sebanyak 19.435 orang, berasal dari rumah sakit yang tersebar di 43 negara. Data pribadi pasien seperti nama dan alamat pasien dirahasiakan begitu juga dengan data rumah sakit yang terlibat. Data diambil dari database *International Stroke Trial*, dan telah diizinkan untuk digunakan secara publik.

Data berisi detail mengenai kondisi pasien pada saat sebelum perawatan dilakukan, pada saat perawatan dan pada saat pemantauan sesudah perawatan. Pemantauan masa perawatan dilakukan selama 14 hari, dan pemantauan sesudah perawatan dilakukan selama 6 bulan.

Pada saat sebelum perawatan dimulai data yang dikumpulkan adalah data mengenai kondisi pasien, seperti gejala yang di derita, informasi dasar pasien (umur, jenis kelamin), data mengenai hasil tindakan sebelum diagnosis dan data hasil diagnosis (jenis stroke yang diderita pasien) . Data yang direkam selama perawatan adalah data mengenai tindakan medis yang dilakukan terhadap pasien (misalnya pemberian obat atau pembedahan) dan kejadian apa saja yang terjadi selama masa 14 hari (sudah bisa dipulangkan dari rumah sakit, meninggal, komplikasi, kambuh dan lain – lain) serta kondisi tubuh pasien . Data setelah perawatan merupakan hasil pemantauan selama 6 bulan terhadap pasien, hasil akhir yang direkam merupakan kondisi terakhir pasien pada masa

pemantauan 6 bulan, yaitu meninggal (*dead*), sembuh (*recovered*), belum sembuh (*not recovered*), dependen (*dependen*), dan *missing status* untuk pasien yang tidak diketahui kondisi akhirnya.

4.2. Tahapan Data Preprocessing

Proses pertama dalam tahapan ini data pasien stroke akan dibagi menjadi dua bagian, tabel 4-1 dibawah ini akan menjelaskan detail dari isi atribut data pasien.

Tabel 4-1 Detail data pasien stroke

Data awal pasien	
HOSPNUM	Kode rumah sakit
RDELAY	Rentang waktu terjadinya stroke dengan perawatan
RCONSC	Tingkat kesadaran pasien (F - <i>fully alert</i> / sadar , D - <i>drowsy</i> / setengah sadar, U - <i>unconscious</i> / tidak sadar)
SEX	Jenis Kelamin (M = <i>male</i> /laki - laki ; F = <i>female</i> /perempuan
AGE	Umur pasien (dibulatkan keatas)
RSLEEP	Gejala nampak pada saat pasien tidur (Y/N)
RATRIAL	<i>Atrial fibrillation</i> / detak jantung abnormal (Y/N)
RCT	CT scan sebelum perawatan (Y/N)
RVISINF	Terlihat infark pada hasil CT scan (Y/N)
RHEP24	Pemberian heparin dalam jangka waktu 24 jam sebelum perawatan (Y/N)
RASP3	pemberian aspirin dalam waktu 3 hari sebelum perawatan (Y/N)
RSBP	Tekanan darah sistolik (mmHg)
RDEF1	defisit pada wajah (Y/N/C=can't assess)
RDEF2	defisit pada lengan/tangan (Y/N/C=can't assess)
RDEF3	defisit pada kaki (Y/N/C=can't assess)
RDEF4	Dysphasia / kelainan pemahaman bahasa (Y/N/C=can't assess)

RDEF5	Hemianopia / pengurangan kemampuan penglihatan (Y/N/C=can't assess)
RDEF6	Visuospatial disorder / kelainan kemampuan pengenalan tempat / lingkungan (Y/N/C=can't assess)
RDEF7	Tanda pendarahan otak (Y/N/C=can't assess)
RDEF8	defisit lain (Y/N/C=can't assess)
STYPE	subtipe stroke (TACS/PACS/POCS/LACS/OTH=other)
RDATE	tahun dan bulan perawatan(yyyy-mm)
HOURLOCAL	Jam
MINLOCAL	Menit
DAYLOCAL	Hari ; 1 - Sunday, 2-Monday, 3-Tuesday, 4-Wednesday, 5-Thursday, 6-Friday, 7-Saturday
RXASP	Alokasi trial aspirin (Y/N)
RXHEP	Alokasi trial heparin (M/L/N).
Data yang dikumpulkan selama 14 hari atau masa perawatan	
DASP14	Pemberian aspirin selama 14 hari/ meninggal/ keluar (Y/N/U=unknown)
DASPLT	Keluar dari perawatan dengan konsumsi aspirin dalam jangka waktu panjang (Y/N/U=unknown)
DLH14	pemberian heparin dosis kecil selama perawatan/meninggal/keluar(Y/N/U=unknown)
DMH14	pemberian heparin dosis medium selama perawatan/meninggal/keluar(Y/N/U=unknown)
ONDRUG	Masa perawatan (hari)
DSCH	Non trial subcutaneous heparin / heparin bawah kulit (Y/N/U=unknown)
DIVH	Non trial intravenous heparin / heparin melalui pembuluh darah (Y/N/U=unknown)
DAP	pemberian obat non trial antiplatelet (Y/N/U=unknown)
DOAC	pemberian antikoagulan jenis lain (Y/N/U=unknown)
DGORM	Glycerol/manitol (Y/N/U=unknown)
DSTER	Steroids (Y/N/U=unknown)
DCAA	kalsium antagonis (Y/N/U=unknown)

DHAEMD	Haemodilution / penambahan volume plasma darah (Y/N/U=unknown)
DCAREND	Operasi Carotid (Y/N/U=unknown)
DTHROMB	Thrombolysis / pemecahan sumbatan darah (Y/N/U=unknown)
DMAJNCH	Pendarahan non-cerebral mayor (Y/N/U=unknown)
DMAJNCHD	Keterangan tanggal
DMAJNCHX	Keterangan DMAJNCH
DSIDE	Efek samping lain (Y/N/U=unknown)
DSIDED	Keterangan tanggal
DSIDEX	Keterangan DSIDE

DIAGNOSIS

DDIAGISC	Ischaemic stroke (Y/N/U=unknown)
DDIAGHA	Haemorrhagic stroke (Y/N/U=unknown)
DDIAGUN	Indeterminate stroke (Y/N/U=unknown)
DNOSTRK	Not a stroke (Y/N/U=unknown)
DNOSTRKX	Keterangan

Stroke yang kambuh dalam masa perawatan

DRSISC	Ischaemic stroke kambuh (Y/N/U=unknown)
DRSISCD	Keterangan tanggal
DRSH	Haemorrhagic stroke kambuh (Y/N/U=unknown)
DRSHD	Keterangan tanggal
DRSUNK	Tipe tidak diketahui (Y/N/U=unknown)
DRSUNKD	Keterangan tanggal

Other events within 14 days

DPE	Pulmonary embolism; (Y/N/U=unknown)
DPED	Keterangan tanggal
DALIVE	Selesai perawatan dalam keadaan hidup (Y/N/U=unknown)
DALIVED	Keterangan tanggal

DPLACE	Tempat tinggal setelah perawatan (A-Home/B-Relatives home/C-Residential care/D-Nursing home/E-Other hospital departments/U-Unknown)
DDEAD	Meninggal dalam masa perawatan (Y/N/U=unknown)
DDEADD	Keterangan
DDEADC	Penyebab kematian (1-Initial stroke/2-Recurrent stroke (ischaemic or unknown)/3-Recurrent stroke (haemorrhagic)/4-Pneumonia/5-Coronary heart disease/6-Pulmonary embolism/7-Other vascular or unknown/8-Non-vascular/0-unknown)
DDEADX	Keterangan

Data yang dikumpulkan selama 6 bulan pemantauan

FDEAD	Meninggal dalam masa pemantauan 6 bulan (Y/N/U=unknown)
FLASTD	Waktu terakhir melakukan kontak dengan pasien (hari)
FDEADD	selang waktu antara perawatan dan waktu meninggal
FDEADC	Penyebab kematian (1-Initial stroke/2-Recurrent stroke (ischaemic or unknown)/3-Recurrent stroke (haemorrhagic)/4-Pneumonia/5-Coronary heart disease/6-Pulmonary embolism/7-Other vascular or unknown/8-Non-vascular/0-unknown)
FDEADX	Keterangan
FRECOVER	Pasien sembuh (Y/N/U=unknown)
FDENNIS	Pasien cacat/ harus tergantung dengan orang lain selama pemantauan (Y/N/U=unknown)
FPLACE	Tempat tinggal pasien selama 6 bulan pemantauan (A-Home/B-Relatives home/C-Residential care/D-Nursing home/E-Other hospital departments/U-Unknown)
FAP	pasien mengkonsumsi obat antiplatelet selama pemantauan (Y/N/U=unknown)
FOAC	pasien mengkonsumsi obat anticoagulant jenis oral pada masa pemantauan(Y/N/U=unknown)

Data lain

FU1_RECD	Jumlah hari dari awal pemantauan sampai selesai 14 hari perawatan
FU2_DONE	Jumlah hari pemantauan setelah perawatan

COUNTRY	Kode negara tempat perawatan pasien
CNTRYNUM	Kode negara
FU1_COMP	Total hari perawatan dan pemantauan
NCCODE	Coding of compliance
CMPLASP	Compliant aspirin (N/Y)
CMPLHEP	Compliant heparin (N/Y)
ID	Indikator variabel kemaian (1 = died; 0 = did not die)
TD	Time of death (jika pasien meninggal, jika pasien belum meninggal variabel ini berisi jumlah hari perawatan dan pemantauan)
EXPDD	Prediksi kemungkinan pasien meninggal / dependent pada waktu 6 bulan
EXPD6	Prediksi kemungkinan kematian dalam waktu 6 bulan
EXPD14	Prediksi kemungkinan kematian dalam waktu 14 hari perawatan
SET14D	Pasien diketahui hidup atau meninggal dalam waktu 14 hari perawatan (1 = Yes, 0 = No)
ID14	Indikasi kematian dalam waktu 14 hari (1 = Yes, 0 = No)
OCCODE	Hasil dari pemantauan 6 bulan(1-dead/2-dependent/3-not recovered/4-recovered/0 or 9 - missing status
Atribut keterangan spesifik penyebab kematian pasien	
DEAD1	Stroke diagnosis awal(1 = Yes, 0 = No)
DEAD2	stroke kambuh (ischaemic/unknown stroke) (1 = Yes, 0 = No)
DEAD3	Stroke Haemorrhagic kambuh (1 = Yes, 0 = No)
DEAD4	Pneumonia (1 = Yes, 0 = No)
DEAD5	Penyakit jantung (1 = Yes, 0 = No)
DEAD6	Pulmonary embolism (1 = Yes, 0 = No)
DEAD7	Penyakit vascular lain / unknown (1 = Yes, 0 = No)
DEAD8	Penyakit non vascular (1 = Yes, 0 = No)
H14	Pendarahan otak/stroke haemorrhagic dalam waktu 14 hari perawatan (1 = Yes, 0 = No)
ISC14	ischaemic stroke dalam waktu 14 hari perawatan (1 = Yes, 0 = No)

NK14	Inderetminate stroke dalam waktu 14 hari (1 = Yes, 0 = No)
STRK14	Indikator stroke dalam waktu 14 hari(1 = Yes, 0 = No)
HTI14	haemorrhagic transformation (1 = Yes, 0 = No)
PE14	pulmonary embolism dalam waktu 14 hari (1 = Yes, 0 = No)
DVT14	deep vein trhomboisis dalam waktu 14 hari (1 = Yes, 0 = No)
TRAN14	indikasi pendarahan non cerebral major dalam waktu 14 hari (1 = Yes, 0 = No)
NCB14	indikasi pendarahan non cerebral dalam waktu 14 hari (1 = Yes, 0 = No)

Data diatas akan dibagi menjadi dua bagian, yang pertama merupakan data awal pasien sampai dengan diagnosis, yang kedua merupakan data perawatan pasien selama 14 hari dan pemantauan selama 6 bulan.

4.2.1. Penambahan atau pengurangan atribut

Berikut ini merupakan setail pembagian data menjadi dua bagian:

a. Data Diagnosis

Tabel 4-2 Tabel atribut data diagnosis yang dihilangkan

Atribut	Koefisien kontingensi	P-Value
RXASP	0.011	0.633
HOURLLOCAL	0.093	0.454
MINLOCAL	0.077	0.454
RHEP24	0.02	0.122
RSBP	0.185	0.092
RDATE	0.178	0.092
RASP3	0.081	0.082

Pada tabel 4-2 diatas, terlihat hasil koefisien kontingensi dan nilai p- value. Atribut yang memiliki nilai p-value lebih besar dari 0.05 tidak ikut dimasukkan. Selain pemilihan atribut berdasarkan korelasi, atribut yang memiliki isi atau nilai yang sama sebaiknya digabung untuk mempermudah proses *association rule* mining dan diharapkan prosesnya bisa lebih cepat.

Tabel 4-3 Penggabungan atribut diagnosis

Penggabungan Atribut	
DDIAGISC	DIAGNOSIS
DDIAGHA	
DDIAGUN	
DNOSTRK	
DNOSTRKX	

Atribut diatas digabungkan menjadi satu atribut karena memiliki isi yang sama, yaitu :

- DDIAGISC : Pasien didiagnosis stroke *ischaemic*
- DDIAGHA : Pasien didiagnosis stroke *Haemorrhagic*
- DDIAGUN : Stroke belum bisa ditentukan
- DNOSTRK : Pasien tidak mengidap stroke
- DNOSTRKX : keterangan

Secara garis besar atribut berisi keterangan diagnosis, oleh karena itu atribut dijadikan satu menjadi atribut diagnosis, dimana akan bernilai *Ischaemic* apabila nilai DDIAGISC=Y, *Haemorrhagic* apabila nilai DDIAGHA=Y, *Indeterminate* apabila nilai DDIAGUN=Y, *Not a stroke* apabila nilai DNOSTRK=Y dan selebihnya bernilai *unknown*.

b. *Data Survival*

Tabel dibawah ini berisi atribut data survival yang dihilangkan karena nilai p -value lebih besar dari 0.05

Tabel 4-4 Tabel atribut data survival yang dihilangkan

Atribut	Koefisien Kontingensi	P-Value
DDEADD	1.2	1
DRSUNKD	0.67	0.963
DMAJNCHD	0.695	0.947
DRSISCD	0.541	0.938
DPED	0.701	0.93
FU1_COMP	0.763	0.807
FLASTD	0.837	0.685
DSIDED	0.494	0.557
DALIVED	0.271	0.491
DRSHD	0.77	0.461
H14	0.021	0.266
NCB14	0.021	0.201
FU1_RECD	0.04	0.09
NK14	0.26	0.073
DVT14	0.26	0.073
expd6	0.864	0.071
TRAN14	0.33	0.068
FU2_DONE	0.85	0.067
expdd	0.699	0.06
expd14	0.538	0.058

Tabel 4-5 Atribut data survival yang dihilangkan(2)

Atribut Dihilangkan	Keterangan
DMAJNCHX	
DSIDEX	Berisi Keterangan atau komentar
DDEADX	
FDEADX	
DIED	Isi sudah tercakup dalam atribut OCCODE yang berisi hasil perawatan pasien
DALIVE	
FRECOVER	
FDENNIS	
DDEAD	
FDEAD	
DEAD1	
DEAD2	
DEAD3	
DEAD4	
DEAD5	Isi sudah terkandung dan sama dengan atribut FDEADC & DDEADC yang berisi penyebab kematian pasien. Digabung menjadi atribut DCAUSE
DEAD6	
DEAD7	
DEAD8	

Pada tabel 4-5 atribut dihilangkan karena berisi komentar dan beberapa atribut digabung menjadi satu. Atribut yang digabungkan dipastikan memiliki isi atau nilai yang sama atau nilai yang tercantum sudah dicakup oleh atribut yang lain. Untuk memastikan, contoh data dapat dilihat pada lampiran A.

4.2.2. Pemilihan atribut yang relevan dengan koefisien kontingensi

Pada bagian ini akan dijelaskan atribut yang terpilih untuk dipakai pada proses *association rule mining*.

a. Data Diagnosis

Tabel dibawah ini berisi atribut – atribut yang akan diproses lebih lanjut di dalam data diagnosis. Keterangan atribut tercantum pada tabel 4-1.

Tabel 4-6 Atribut terpilih untuk data diagnosis

Atribut	Koefisien Kontingensi	P-Value
SEX	0.026	0.01
DAYLOCAL	0.047	0.008
RDEF8	0.034	0.004
RSLEEP	0.032	0.001
RDELAY	0.132	0
RCONSC	0.047	0
AGE	0.181	0
RATRIAL	0.039	0
RCT	0.317	0
RVISINF	0.186	0
RDEF1	0.076	0
RDEF2	0.092	0
RDEF3	0.083	0
RDEF4	0.039	0
RDEF5	0.057	0
RDEF6	0.041	0
RDEF7	0.049	0
STYPE	0.069	0

Atribut	Koefisien Kontingensi	P-Value
RXHEP	0.095	0

b. Data *Survival*

Tabel dibawah ini berisi atribut – atribut yang akan diproses lebih lanjut di dalam data diagnosis. Keterangan atribut tercantum pada tabel 4-1.

Tabel 4-7 Atribut terpilih data survival

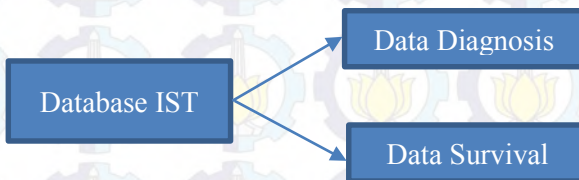
Atribut	Koefisien Kontingensi	P-Value
DHH14	0.098	0.047
DIAGNOSIS	0.579	0
DASP14	0.492	0
DASPLT	0.402	0
DLH14	0.444	0
DMH14	0.395	0
ONDRUG	0.415	0
DSCH	0.418	0
DIVH	0.394	0
DAP	0.385	0
DOAC	0.42	0
DGORM	0.328	0
DSTER	0.351	0
DCAA	0.317	0
DHAEMD	0.325	0
DCAREND	0.31	0
DTHROMB	0.394	0
DMAJNCH	0.432	0

Atribut	Koefisien Kontingensi	P-Value
DSIDE	0.407	0
DRSISC	0.578	0
DRSH	0.577	0
DRSUNK	0.586	0
DPE	0.579	0
DPLACE	0.377	0
DCAUSE	0.709	0
FPLACE	0.364	0
FAP	0.092	0
FOAC	0.074	0
COUNTRY	0.32	0
NCCODE	0.622	0
CMPLASP	0.068	0
CMPLHEP	0.067	0
TD	0.81	0
ISC14	0.129	0
STRK14	0.069	0
HTI14	0.067	0
PE14	0.388	0
SET14D	0.404	0
ID14	0.123	0

4.2.3. Gambaran hirarki data

Data yang diambil berasal dari *International Stroke Trial (IST) database*. Lalu dalam tahapan ini data dibagi menjadi dua, yang pertama adalah data diagnosis dengan atribut hasil “Diagnosis” dan data survival dengan atribut hasil “OCCODE”.

Penggalian rule pada data diagnosis perlu dilakukan secara terpisah karena atribut yang ada di dalam data tersebut hanya memiliki korelasi kuat terhadap atribut hasil “Diagnosis” dan tidak berkorelasi dengan atribut hasil “OCCODE” pada data survival. Selain itu atribut hasil “Diagnosis” merupakan input dari atribut hasil “OCCODE” yang menjelaskan tentang kondisi akhir pasien, sehingga jika diketahui kondisi akhir pasien dapat dilacak kembali diagnosis jenis stroke-nya (jenis stroke tercantum dalam atribut hasil “Diagnosis”) dan faktor – faktor yang mempengaruhi diagnosis jenis stroke tersebut juga dapat dilacak kembali dengan mudah . Gambar 4-1 menunjukkan hirarki data setelah data dibagi menjadi dua bagian.



Gambar 4-1 Hirarki data secara umum

Data berisi banyak atribut, dan di dalam satu atribut terdapat beberapa kelas atau kategori, detail dari atribut dan kelas pada masing – masing data terdapat pada tabel 4-8.

Tabel 4-8 Detail hirarki data

Database	Data	Atribut	Kelas
IST	Diagnosis	SEX	F,M
IST	Diagnosis	DAYLOCAL	Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday
IST	Diagnosis	RDEF8	Y,N,C
IST	Diagnosis	RSLEEP	Y,N
IST	Diagnosis	RDELAY	A,B
IST	Diagnosis	RCONSC	F,D,U
IST	Diagnosis	AGE	A,B,C,D,E,F
IST	Diagnosis	RATRIAL	Y,N
IST	Diagnosis	RCT	Y,N
IST	Diagnosis	RVISINF	Y,N
IST	Diagnosis	RDEF1	Y,N,C
IST	Diagnosis	RDEF2	Y,N,C
IST	Diagnosis	RDEF3	Y,N,C
IST	Diagnosis	RDEF4	Y,N,C
IST	Diagnosis	RDEF5	Y,N,C
IST	Diagnosis	RDEF6	Y,N,C
IST	Diagnosis	RDEF7	Y,N,C
IST	Diagnosis	STYPE	TACS, PACS, POCS, LACS, OTH
IST	Diagnosis	RXHEP	M,L,N
IST	Diagnosis	DIAGNOSIS (Atribut hasil)	<i>Ischaemic, Haemorrhagic, Indeterminate, Not a stroke, unknown</i>
IST	Survival	DHH14	Y,N,U

Database	Data	Atribut	Kelas
IST	Survival	DIAGNOSIS	<i>Ischaemic, Haemorrhagic, Indeterminate, Not a stroke, unknown</i>
IST	Survival	DASP14	Y,N,U
IST	Survival	DASPLT	Y,N,U
IST	Survival	DLH14	Y,N,U
IST	Survival	DMH14	Y,N,U
IST	Survival	ONDRUG	1,2,3,4,5,6,7,8,9,10,11,12,13,14
IST	Survival	DSCH	Y,N,U
IST	Survival	DIVH	Y,N,U
IST	Survival	DAP	Y,N,U
IST	Survival	DOAC	Y,N,U
IST	Survival	DGORM	Y,N,U
IST	Survival	DSTER	Y,N,U
IST	Survival	DCAA	Y,N,U
IST	Survival	DHAEMD	Y,N,U
IST	Survival	DCAREND	Y,N,U
IST	Survival	DTHROMB	Y,N,U
IST	Survival	DMAJNCH	Y,N,U
IST	Survival	DSIDE	Y,N,U
IST	Survival	DRSISC	Y,N,U
IST	Survival	DRSH	Y,N,U
IST	Survival	DRSUNK	Y,N,U
IST	Survival	DPE	Y,N,U

Database	Data	Atribut	Kelas
IST	Survival	DPLACE	A,B,C,D,E,U
IST	Survival	DCAUSE	Initial stroke, Recurrent stroke (ischaemic or unknown), Recurrent stroke (haemorrhagic), Pneumonia, Coronary heart disease, Pulmonary embolism, Other vascular or unknown, Non-vascular,unknown
IST	Survival	FPLACE	A,B,C,D,E,U
IST	Survival	FAP	Y,N,U
IST	Survival	FOAC	Y,N,U
IST	Survival	COUNTRY	Nama negara yang terlibat
IST	Survival	NCCODE	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15
IST	Survival	CMPLASP	Y,N
IST	Survival	CMPLHEP	Y,N
IST	Survival	TD	A,B,C,D,E,F,G,H,I,J
IST	Survival	ISC14	Y,N
IST	Survival	STRK14	Y,N
IST	Survival	HTI14	Y,N
IST	Survival	PE14	Y,N
IST	Survival	SET14D	Y,N
IST	Survival	ID14	Y,N

4.2.4. Konversi atribut bernilai kontinyu menjadi atribut diskrit

Untuk melakukan proses *association rule* mining, atribut harus berisi nilai diskrit atau kategorikal, untuk itu

atribut bernilai kontinyu akan dikonversi menjadi atribut bernilai kategorikal.

Dalam tugas akhir ini digunakan WEKA untuk *data preprocessing*. Caranya dengan menggunakan filter discretization untuk mengubah nilai atribut kontinyu menjadi kategorikal. Contohnya pada atribut AGE, atribut ini berisi nilai yang bersifat kontinyu (gambar), setelah itu nilainya dikonversi dengan filter WEKA dan terbagi menjadi 6 kategori atau kelas.

RCONSC	SEX	AGE
D	M	69
F	M	76
F	F	71
F	M	81
F	M	78
F	M	54
F	F	77
F	M	23
F	M	47
F	M	81
D	M	48
F	F	45
D	F	83
F	F	86
F	M	50

Gambar 4-2 AGE sebelum dikonversi

Selected attribute		
Name: AGE		Type: Nominal
Missing: 0 (0%)	Distinct: 6	Unique: 0 (0%)
No.	Label	Count
1	'(-inf-52.5]'	1368
2	'(52.5-64.5]'	3270
3	'(64.5-74.5]'	6005
4	'(74.5-82.5]'	5362
5	'(82.5-89.5]'	2940
6	'(89.5-inf)'	490

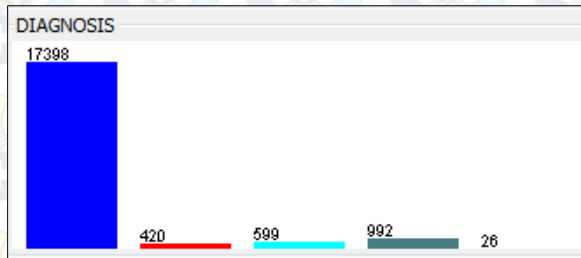
Gambar 4-3 Atribut AGE setelah dikonversi bernilai diskrit

4.2.5. Pembagian data *training* dan data *testing*

Untuk pembagian data *training* dan *testing* akan digunakan filter resample. Data *training* akan dialokasikan sebesar 70% dan 30% untuk data *testing*.

Pada saat melakukan pemecahan data menjadi data *training* dan *testing*, perlu diperhatikan distribusi kelas dalam suatu atribut. Sebaiknya distribusi kelas di dalam atribut pada data *training* sama dengan data *testing*.

Gambar dibawah ini akan menunjukkan distribusi kelas pada atribut Diagnosis dan OCCODE, dimana atribut diagnosis merupakan atribut hasil dari data diagnosis dan atribut OCCODE merupakan atribut hasil dari data survival. Atribut diagnosis berisi diagnosis penyakit stroke tiap pasien, atribut OCCODE berisi hasil dari perawatan pasien selama 14 hari dan pemantauan selama 6 bulan.

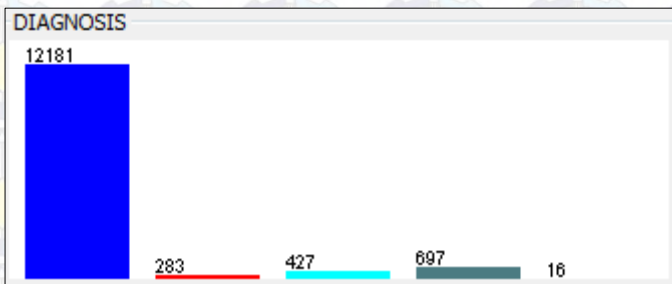


Gambar 4-4 Diagram distribusi kelas atribut diagnosis

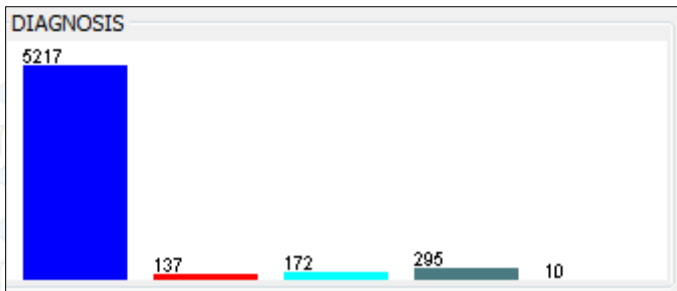
Selected attribute		
Name: DIAGNOSIS		Type: Nominal
Missing: 0 (0%)	Distinct: 5	Unique: 0 (0%)
N	Label	Count
1	Ischaemic	17398
2	Not a stroke	420
3	Haemorrhagic	599
4	Indeterminate	992
5	unknown	26

Gambar 4-5 Keterangan atribut diagnosis

Pada gambar 4-4, warna biru mewakili Ischaemic, merah Not a stroke, biru muda Haemorrhagic, abu – abu indeterminate dan yang terakhir adalah unknown. Gambar diatas menunjukkan distribusi kelas atribut diagnosis sebelum dipecah menjadi data *training* dan *testing*.

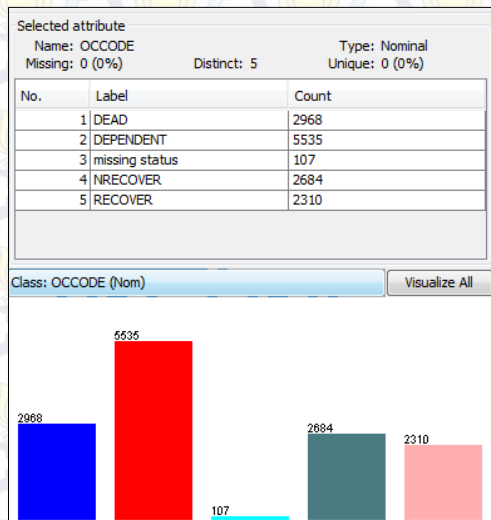


Gambar 4-6 Atribut diagnosis pada data *training*

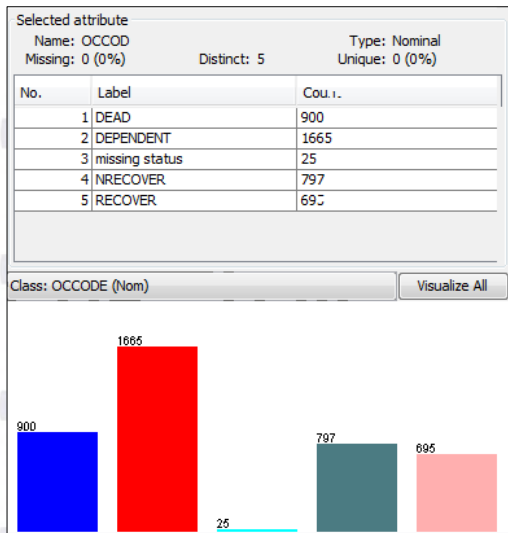


Gambar 4-7 Atribut diagnosis pada data *testing*

Pada gambar 4-6 dan 4-7 terlihat bahwa distribusi kelas dalam atribut diagnosis tidak berubah. Gambar 4-8 dan 4-9 akan menunjukkan distribusi kelas pada atribut OCCODE pada, data *training* dan data *testing*.



Gambar 4-8 OCCODE data *training*



Gambar 4-9 OCCODE data testing

Pada gambar 4-8 dan 4-9 terlihat distribusi kelas pada data *training* dan *testing* tidak berubah.

4.3. Penyiapan file sesuai dengan format

Setelah data dipecah menjadi data *training* dan *testing*, akan disiapkan data sesuai dengan format yang dapat dibaca oleh aplikasi RStudio. Aplikasi ini memiliki banyak *package* yang dapat diinstall sendiri oleh user. Salah satu *package* yang dapat dipakai untuk membaca file adalah *package foreign*. Berfungsi untuk membaca file dari aplikasi lain seperti excel, minitab, spss dan WEKA. Karena sebelumnya data di praproses menggunakan WEKA, akan dipakai file dengan format *.arff* untuk diproses di dalam RStudio.

BAB V

TAHAP PELATIHAN DAN PENGUJIAN

Pada bab ini akan dijelaskan mengenai proses *association rule mining*. Prosesnya terbagi menjadi dua tahapan, yaitu tahap penggalian *rule* pada data *training* dan tahap pengujian *rule* pada tahap *testing*.

5.1. Penggalian Rule

Untuk melakukan tahap ini digunakan aplikasi RStudio, penggalian *rule* menggunakan metode *association rule mining* dengan algoritma apriori. Berikut ini adalah potongan kode program yang ditulis di RStudio

```
#arules dan arulesviz untuk menjalankan association rule mining
#foreign untuk mengakses file berasal dari luar format rstudio (.arff)
#load colorspace library
library(arules)
library(arulesviz)
library(foreign)
library(colorspace)
```

Gambar 5-1 Potongan kode load library

Gambar 5-1 diatas merupakan kode untuk memanggil *library* , *arules* berfungsi untuk melakukan proses *association rule mining* dengan algoritma apriori. *ArulesViz* berfungsi untuk melakukan visualisasi pada output apriori. *Foreign* berfungsi untuk membaca file yang memiliki format dari luar RStudio, seperti file dari minitab, spss, WEKA, excel dan lain sebagainya.

```
#load data (pilih data latih / training)
survival<-read.arff(file="survivaltraining.arff")
```

Gambar 5-2 Potongan kode untuk load data

Setelah *load* data, RStudio sudah menyimpan data di dalam obyek, pada gambar diatas data disimpan di dalam obyek *survival*, setelah itu kita dapat melanjutkan dengan proses *association rule* mining menggunakan algoritma apriori. Potongan kode dapat dilihat pada gambar 5-3.

```
#menjalankan algoritma apriori
#parameter minsup dan mincof dapat disesuaikan
#rhs dapat disesuaikan
rulesurvival<-apriori(survival, parameter= list(support=0.109,confidence=0.130),
  appearance=list(rhs= c("OCCODE=DEAD"),default="lhs"))
```

Gambar 5-3 Penerapan algoritma apriori

Dalam gambar tersebut, harus dimasukkan parameter minimum *support* dan minimum *confidence*. Minimum *support* menggambarkan proporsi *rule* terhadap keseluruhan transaksi. Misalnya saja minimum *support* 0.5, berarti *rule* tersebut dapat ditemukan pada 50% data keseluruhan. Selain menetapkan nilai minsup dan mincof, kita juga dapat memilih bagian rhs (right hand side) atau bagian consequent rule yang akan dikeluarkan, pada gambar diatas consequent yang dipilih adalah rule yang memiliki hasil pasien berstatus meninggal (*dead*).

Jika dilihat dari bab VI bagian pembagian data *training* dan *testing*, data yang ada tidak memiliki penyebaran kelas yang rata. Pada data diagnosis kelas yang mendominasi adalah stroke ischaemic.

Selected attribute		
Name: DIAGNOSIS		
Missing: 0 (0%)		Type: Nominal
Distinct: 5		Unique: 0 (0%)
No.	Label	Count
1	Ischaemic	12181
2	Not a stroke	283
3	Haemorrhagic	427
4	Indeterminate	697
5	unknown	16

Gambar 5-4 proporsi data *training* diagnosis

Begitu juga dengan data survival, penyebaran distribusi kelas pada atribut OCCODE juga kurang merata. Tabel dibawah ini merangkum prosentase distribusi masing – masing kelas terhadap atributnya.

Tabel 5-1 Prosentase kelas terhadap keseluruhan jumlah data dalam atribut

PROPORSI DATA			Minsup	Mincof
<i>TRAINING</i>	Jumlah	%		
Ischaemic	12181	89.540	0.448	0.537
Haemorrhagic	283	2.080	0.010	0.012
Indeterminate	427	3.139	0.016	0.019
Not a stroke	697	5.123	0.026	0.031
unknown	16	0.118	0.001	0.001
Total	13604	100.000	0.500	0.600
Dead	2068	21.718	0.109	0.130
Recover	1615	16.961	0.085	0.102
Not recover	1887	19.817	0.099	0.119
Dependent	3870	40.643	0.203	0.244
missing status	82	0.861	0.004	0.005
Total	9522	100.000	0.500	0.600

Jika dilihat pada tabel diatas, tidak bisa langsung ditentukan nilai minimum *confidence* dan minimum *support* untuk perhitungan *association rule mining*. Bisa saja nilai minimum *support* langsung ditentukan, namun jika nilai minimum *support* terlampau besar, kelas yang memiliki prosesntase kecil tidak akan terjangkau, dengan kata lain *rule* nya tidak bisa keluar.

Oleh karena itu digunakan parameter *minsup* dan *minconf* pada tabel diatas untuk perhitungan apriori pada RStudio. Nilai minimum *support* ditentukan paling tidak *rule* tersebut ada di dalam 50% data. Misalnya pada atribut OCCODE dengan kelas Recover, kelas recover memiliki prosesntase sebesar 16.9% dari keseluruhan data, maka nilai *minsup*nya adalah 50% dari 16.9 yaitu 0.085. nilai *minconf* sebesar 60% dari 16.9 yaitu 0.102. Nilai 60% diambil karena *confidence* sebenarnya menggambarkan kekuatan keeratan di dalam *rule*, menurut tabel korelasi 60% merupakan nilai korelasi kuat. Oleh karena itu ditentukan nilai *minconf* sebesar 60% dari transaksi kelas tersebut.

Penentuan minimum *confidence* dan minimum *support* diatas bertujuan agar semua kelas memiliki output pada saat penggalian *rule*. Jika nilai minimum *support* dan minimum *confidence* ditentukan secara umum, maka yang akan memiliki output *rule* hanya kelas yang dominan saja atau yang memenuhi nilai minimum *support/confidence* yang ditentukan tersebut.

```
#sort rule by lift
rulessurvival.sorted<-sort(rulessurvival, by="lift")
#prune redundant rules
subset.matrix<-is.subset(rulessurvival.sorted, rulessurvival.sorted)
subset.matrix[lower.tri(subset.matrix,diag=T)]<-NA
redundant<-colsums(subset.matrix,na.rm=T)
#menghapus redundant rules
rulessurvival.pruned<-rulessurvival.sorted[!redundant]
#menampilkan
inspect(rulessurvival.pruned)
```

Gambar 5-5 Pruning rule

Setelah melakukan penggalian *rule* dengan algoritma apriori, tentunya kita ingin *rule* yang ada tidak ada yang redundan

atau duplikat. Gambar 5-5 merupakan potongan kode untuk melakukan *pruning*, pengukur yang akan digunakan sebagai dasar *pruning* adalah nilai *lift*. Apabila *super rule* tidak memiliki nilai *lift* yang lebih besar dari *rule* dibawahnya, maka *super rule* tersebut akan dipangkas karena tidak memiliki derajat kepentingan lebih tinggi dan tidak memberikan pengetahuan baru.

Potongan kode dibawah ini berfungsi untuk menghitung nilai *coverage* dari setiap *rule*. Kita tidak perlu melakukan perhitungan *support*, *confidence* dan *lift* dari masing – masing *rule* karena nilai tersebut sudah terhitung pada proses implementasi algoritma *apriori* dan *pruning*.

```
#menghitung coverage rule
quality(rulesurvival) <- cbind(quality(rulesurvival),
                               coverage = coverage(rulesurvival))
```

Gambar 5-6 Menghitung *coverage*

5.2. Pengujian *Rule*

Proses selanjutnya setelah penggalian *rule* dari data *training* adalah *testing*. *Testing* bertujuan untuk mengetahui seberapa akurat masing – masing *rule* apabila diterapkan.

Hasil atau output *rule* dari proses sebelumnya dapat dilihat secara lengkap pada lampiran A2 dan A3. Sistem dari pengujian ini adalah setiap *rule* akan dicocokkan dengan data *testing*, apakah *rule* tersebut mendeteksi kelasnya dengan benar atau tidak, jika bernilai *rule* mendapat nilai *true* dan jika salah akan mendapatkan nilai *false*. Akurasi akan dihitung dari perbandingan antara jumlah nilai *true* dengan jumlah nilai *true*+*false* (keseluruhan data yang memiliki kelas *rule* tersebut).

Untuk melakukan hal diatas akan digunakan alat bantu yang dibuat secara manual, alat bantu ini dibangun menggunakan bahasa pemrograman java. Dibangun menggunakan *Interactive Development Environment* (IDE) Eclipse Luna. Berikut ini merupakan potongan kode yang ada di dalam sistem sederhana tersebut.

Potongan kode pada gambar 5-7 berfungsi untuk memasukkan rule dari hasil *pruning*. Jika ada 10 rule maka akan ada 10 kondisi di dalam kode program. Rule ini akan dicocokkan dengan data *testing* yang disimpan di dalam format excel (.xls).

```
package com.main;

public class recover {

    public String[] statKlas(Double DSCH,
        Double DIVH, Double DAP, Double DOAC, Double DGORM,
        Double DCAA, Double DHAEMD, Double DCAREND, Double DTHROMB,
        Double DMAJNCH, Double DCAUSE, Double CMPLASP, Double CMPLHEP,
    )
    {
        String[] has = new String[10];
        for (int a=0; a < 10; a++){
            has[a] = "False";
        }
        if (DTHROMB==1&&DMAJNCH==1&&DCAUSE==9&&CMPLHEP==2){
            has[0] = "T";
        }
        if (DTHROMB==1&&DCAUSE==9&&CMPLHEP==2){
            has[1] = "T";
        }
        if (DTHROMB==1&&DCAUSE==9&&CMPLASP==2){
            has[2] = "T";
        }
        if (DSCH==1&&DTHROMB==1&&DMAJNCH==1&&DCAUSE==9){
            has[3] = "T";
        }
        if (DSCH==1&&DMAJNCH==1&&DCAUSE==9){
            has[4] = "T";
        }
        if (DSCH==1&&DTHROMB==1&&DCAUSE==9){
            has[5] = "T";
        }
        if (DMAJNCH==1&&DCAUSE==9&&CMPLHEP==2){
            has[6] = "T";
        }
        if (DCAUSE==9&&CMPLHEP==2){
            has[7] = "T";
        }
        if (DSCH==1&&DCAUSE==9){
            has[8] = "T";
        }
        if (DMAJNCH==1&&DCAUSE==9&&CMPLASP==2){
            has[9] = "T";
        }
    }
}
```

Gambar 5-7 Potongan kode memasukkan rule

Potongan kode pada gambar 5-8 berfungsi untuk memilih file input (data *testing*) dan memilih file tempat penyimpanan *output*. Lalu akan dilakukan pencocokkan dengan data *testing* setelah memanggil file data *testing* dan kelas tempat *rule* dimasukkan.

```

public static void main(String[] args) throws IOException {
    MainRun test = new MainRun();
    test.setInputFile("D:/BISMILLAH/Stroke/DATA TESTING/nrecover.xls");
    File output = new File("D:/BISMILLAH/Stroke/DATA TESTING/hasilnrecover.xls");
    test.setData(test.baca());
    int bar = test.data[0].length;

    String[][] has = new String[16][bar];
    String[] hasSem = null;
    recover kp = new recover();
    for(int x=0; x<bar; x++){
        hasSem = kp.statKlas(test.data[0][x],test.data[1][x],test.data[2][x],
            test.data[3][x],test.data[4][x],test.data[5][x],test.data[6][x],
            test.data[7][x],test.data[8][x],test.data[9][x],test.data[10][x],
            test.data[11][x],test.data[12][x],test.data[13][x]);

        for(int y=0; y < 16; y++){
            has[y][x] = hasSem[y]; //memasukkan hasil ke tiap baris
        }
    }

    test.setHasil(has);
    test.tulis(output, test.hasil);
    system.out.println("sukses lihat hasil");
}

```

Gambar 5-8 Potongan kode *testing*

BABVI

EVALUASI PERFORMA DAN ANALISIS HASIL

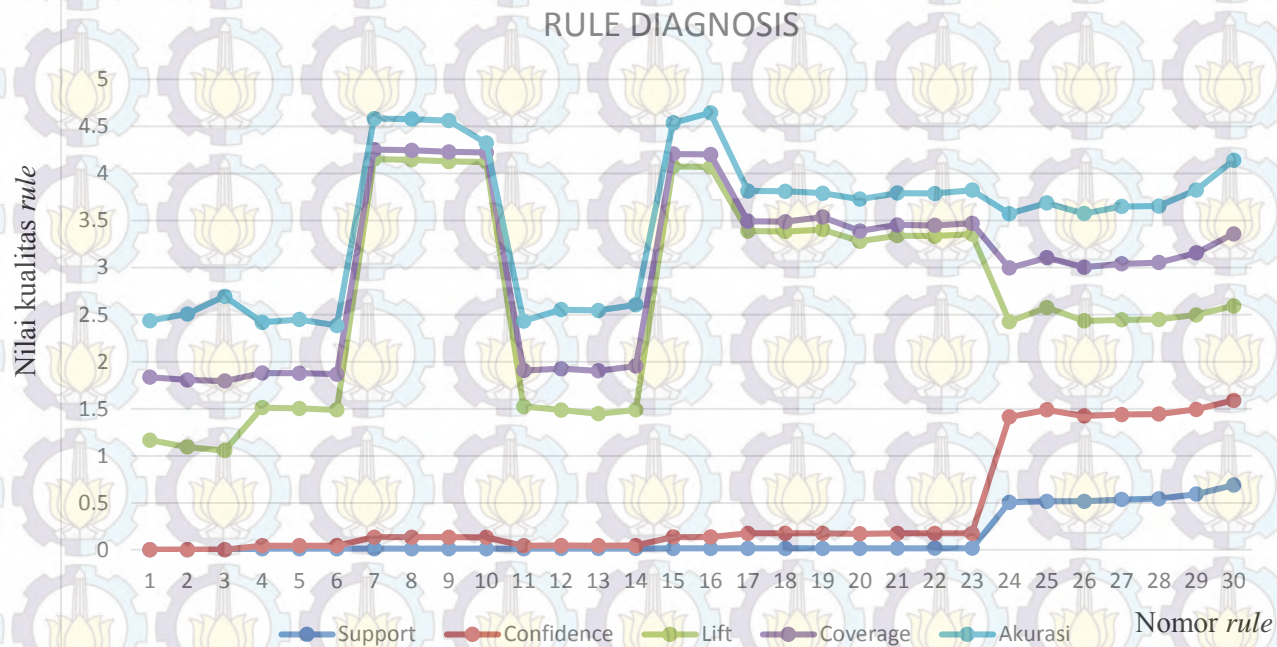
Dalam bab ini akan dianalisis pengaruh perubahan minsup dan mincof terhadap kualitas *rule*. Selain itu *rule* akan dianalisis sehingga diketahui faktor – faktor yang berpengaruh pada masing – masing data diagnosis dan data survival.

6.1. Pengaruh perubahan minimum *support* dan minimum *confidence* pada kualitas *rule*

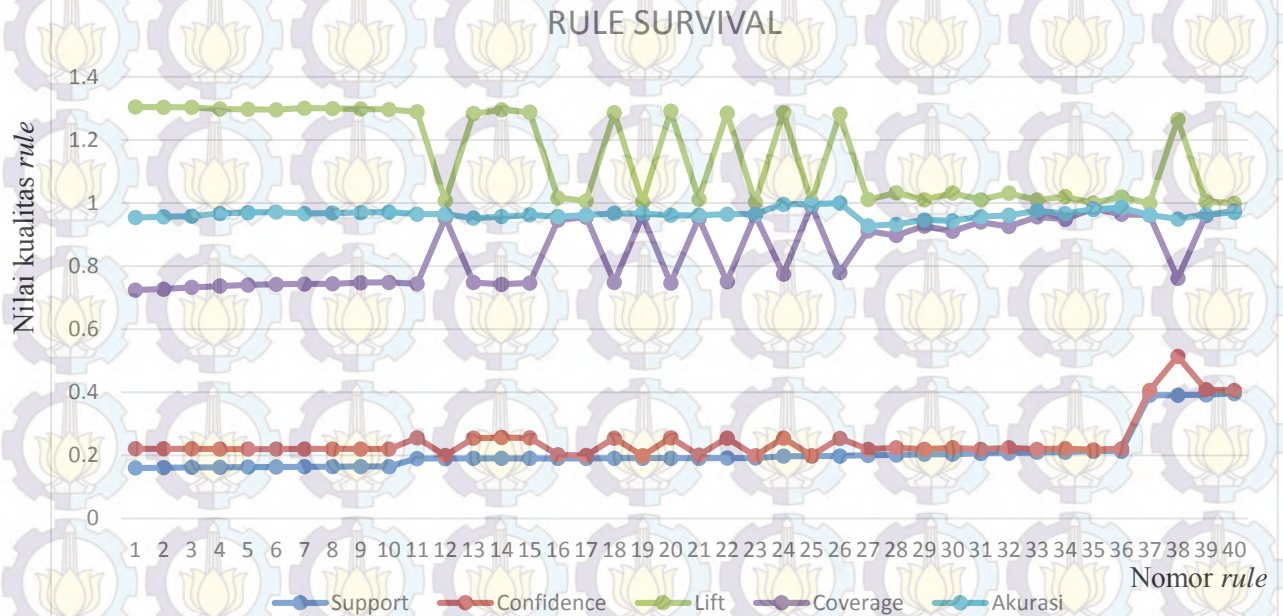
Pada grafik 6-1 dan 6-2 dibawah ini, menunjukkan pengaruh perubahan minimum *support* dan *confidence* terhadap kualitas *rule* yang dihasilkan. Semakin besar nilai minimum *support* dan minimum *confidence* maka jumlah *rule* yang dihasilkan semakin sedikit.

Pada data diagnosis, fluktuasi nilai kualitas *rule* (*lift*, *coverage* dan akurasi) tidak begitu dipengaruhi oleh nilai *support*, namun pada *rule* nomor 24 terjadi kenaikan nilai *support* yang cukup drastis dan hal ini mempengaruhi nilai *lift*, *coverage* dan akurasi, ketiga nilai tersebut terlihat menurun. Hal ini disebabkan karena semakin besar nilai *support* maka cakupan data semakin kecil sehingga nilai *coverage*, *lift* dan akurasinya juga semakin kecil

Sedangkan nilai *confidence* terlihat memiliki pengaruh terhadap naik turunnya nilai kualitas *rule* (*lift*, *coverage* dan akurasi), misalnya saja pada *rule* nomor 7 sampai dengan 10, kenaikan nilai *confidence* mengakibatkan nilai ketiga kualitas *rule* naik drastis. Namun pada *rule* nomor 24 sampai 30. Terlihat penurunan kualitas karena cakupan data yang terlalu sedikit meskipun nilai *confidencenya* besar.



Gambar 6-1 Grafik pengukur performa *rule* diagnosis



Gambar 6-2 Grafik pengukur performa *rule survival*

Pada data survival perubahan nilai *support* tidak begitu berpengaruh pada fluktuasi nilai kualitas *rule*. Nilai *confidence* memiliki pengaruh yang besar pada fluktuasi nilai *lift* dan *coverage*. Semakin besar nilai *confidence* (korelasi) maka nilai *lift* semakin besar dalam arti *rule* semakin menarik dan penting. Sebaliknya semakin besar nilai *confidence* maka nilai *coverage* semakin kecil karena cakupan *rule* terhadap data juga semakin kecil. Nilai akurasi pada data survival terlihat stabil, hal ini disebabkan oleh didistribusi kelas pada atribut OCCODE lebih merata. Berbeda dengan nilai akurasi yang terdapat pada data diagnosis, nilai akurasi terlihat tidak stabil karena distribusi kelas pada atribut Diagnosis tidak merata (didominasi jenis stroke *ischaemic*)

6.2. Analisis *rule*

6.2.1. Data Diagnosis

Pada bagian ini akan dibahas *rule* yang muncul dari setiap kelas pada atribut Diagnosis.

a. *Ischaemic*

Tabel dibawah ini menunjukkan *rule* yang terbentuk setelah proses *pruning*, pola dibawah menggambarkan ada beberapa atribut yang berkontribusi dalam diagnosis stroke *ischaemic*. Dan nilai akurasi menunjukkan kekuatan *rule* tersebut untuk memprediksi diagnosis pada data *testing*. Daftar lengkap *rule* dapat dilihat pada tabel 6-1.

Tabel 6-1 *Rule Stroke Ischaemic*

No	RULE	Akurasi
1	{RATRIAL=N,RCT=Y} => {DIAGNOSIS=Ischaemic}	58.175%

No	RULE	Akurasi
2	{RDEF1=Y,RDEF2=Y,RDEF7=N} => {DIAGNOSIS=Ischaemic}	56.891%
3	{RCONSC=Y,RDEF2=Y,RDEF7=N}=> {DIAGNOSIS=Ischaemic}	57.658%
4	{RCONSC=F,RDEF6=N} => {DIAGNOSIS=Ischaemic}	60.974%
5	{RDEF1=Y,RDEF3=Y} => {DIAGNOSIS=Ischaemic}	60.073%
6	{RCONSC=F} => {DIAGNOSIS=Ischaemic}	78.072%
7	{RDEF6=N} => {DIAGNOSIS=Ischaemic}	66.763%

- **Rule 1**

Pasien yang tidak memiliki kelainan detak jantung (RATRIAL=N) dan telah melakukan proses CT-scan (RCT=Y) akan terdiagnosis mengidap stroke *ischaemic*. Akurasi rule ini sebesar 58.175% terhadap data *testing*.

- **Rule 2**

Pasien yang terdiagnosis stroke *ischaemic* mengalami gejala defisit pada wajah (RDEF1=Y), defisit pada tangan/lengan (RDEF2=Y) dan tidak memiliki tanda pendarahan pada otak (RDEF7=N). Akurasi *rule* ini sebesar 56,891% terhadap data *testing*.

- **Rule 3**

Pasien yang terdiagnosis stroke *ischaemic* dalam keadaan bangun atau sadar (RCONSC=Y), mengalami gejala defisit pada tangan/lengan

(RDEF2=Y) dan tidak menunjukkan tanda pendarahan pada otak (RDEF7=N). Akurasi rule ini sebesar 57.658% pada data *testing*.

• **Rule 4**

Pasien yang terdiagnosis stroke ischaemic dalam keadaan sangat sadar/fully alert (RCONSC=F) dan tidak memiliki gejala kelainan visuospasial / tidak mengenali lingkungan/tempat (RDEF6=N). Akurasi rule ini sebesar 60.974% pada data *testing*.

• **Rule 5**

Pasien yang terdiagnosis stroke ischaemic mengalami gejala defisit pada wajah (RDEF1=Y) dan defisit pada kaki (RDEF3=Y). Akurasi rule ini sebesar 60.073% terhadap data *testing*.

• **Rule 6**

Pasien yang terdiagnosis stroke *ischaemic* dalam keadaan sangat sadar/ *fully alert* (RCONSC=F) pada saat proses diagnosis. Akurasi *rule* ini sebesar 78.072% pada data *testing*.

• **Rule 7**

Pasien yang terdiagnosis stroke *ischaemic* tidak memiliki gejala kelainan visuospasial / tidak mengenali lingkungan / tempat. Akurasi *rule* ini sebesar 66.763% pada data *testing*.

Faktor – faktor yang berpengaruh terhadap diagnosis stroke *ischaemic* adalah sebagai berikut :

1. Tidak adanya kelainan detak jantung pada pasien (RATRIAL=N). Faktor ini terdapat di dalam satu *rule*.
2. Pasien telah melakukan proses CT-scan (RCT=Y). Faktor ini terdapat di dalam satu *rule*.
3. Pasien mengalami gejala defisit pada wajah (RDEF1=Y). Faktor ini muncul sebanyak dua kali / terdapat di dalam dua *rule*.

4. Pasien mengalami defisit pada tangan/lengan (RDEF2=Y). Faktor ini muncul sebanyak dua kali/ terdapat di dalam dua *rule*.
5. Pasien mengalami defisit pada kaki (RDEF3=Y). Faktor ini muncul sebanyak dua kali/ terdapat dalam dua *rule*.
6. Pasien tidak memiliki tanda pendarahan pada otak (RDEF7=N). Faktor ini muncul sebanyak dua kali / terdapat di dalam dua *rule*.
7. Pasien dalam keadaan sadar (RCONSC=Y). Faktor ini muncul sebanyak satu kali.
8. Pasien dalam keadaan sangat sadar / *fully alert* (RCONSC=F). Faktor ini muncul sebanyak dua kali / terdapat di dalam dua *rule*.
9. Pasien tidak mengalami kelainan visuospasial (RDEF6=N). Faktor ini muncul sebanyak dua kali/terdapat di dalam dua *rule*.

b. *Haemorrhagic*

Di dalam tabel 6-2 terdapat daftar dari *rule* pasien yang terdiagnosis stroke jenis *haemorrhagic*. Akurasi dari *rule* kurang baik, karena pada saat penggalian *rule*, minimum *support* nilainya sangat kecil karena proporsi data juga sangat kecil.

Tabel 6-2 *Rule* stroke haemorrhagic

No	<i>RULE</i>	Akurasi
1	{RSLEEP=N,RATRIAL=N,RCT=N,RVISINF=N,RDEF1=Y,RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	33.140%

No	RULE	Akurasi
2	{RSLEEP=N,RATRIAL=N,RCT=N,RDEF1=Y,RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	33.140%
3	{RSLEEP=N,RATRIAL=N,RCT=N,RVISINF=N,RDEF1=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	33.140%
4	{RSLEEP=N,RATRIAL=N,RCT=N,RDEF1=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	9.884%
5	{RSLEEP=N,RATRIAL=N,RCT=N,RVISINF=N,RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	33.140%
6	{RSLEEP=N,RATRIAL=N,RCT=N,RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	44.186%

- **Rule 1**

Pasien yang terdiagnosis stroke *haemorrhagic* tidak memiliki gejala yang tampak pada saat pasien tertidur (RSLEEP=N), tidak memiliki kelainan detak jantung (RATRIAL=N). Tidak melakukan proses CT-scan (RCT=N) dan tidak terlihat infark pada bagian kepala (RVISINF=N). Gejala yang tampak adalah defisit pada wajah (RDEF1=Y), defisit pada lengan/tangan (RDEF2=Y) dan defisit pada kaki (RDEF3=Y). Akurasi *rule* ini sebesar 33.140% terhadap data *testing*.

- **Rule 2**

Pasien yang terdiagnosis stroke *haemorrhagic* tidak memiliki gejala yang tampak pada saat pasien tertidur (RSLEEP=N), tidak memiliki kelainan detak jantung (RATRIAL=N). Tidak melakukan proses CT-scan (RCT=N). Gejala yang tampak

adalah defisit pada wajah (RDEF1=Y), defisit pada lengan/tangan (RDEF2=Y) dan defisit pada kaki (RDEF3=Y). Akurasi *rule* ini sebesar 33.140% terhadap data *testing*.

- **Rule 3**

Pasien yang terdiagnosis stroke *haemorrhagic* tidak memiliki gejala yang tampak pada saat pasien tertidur (RSLEEP=N), tidak memiliki kelainan detak jantung (RATRIAL=N). Tidak melakukan proses CT-scan (RCT=N) dan tidak terlihat infark pada bagian kepala (RVISINF=N). Gejala yang tampak adalah defisit pada wajah (RDEF1=Y) dan defisit pada kaki (RDEF3=Y). Akurasi *rule* ini sebesar 33.140% terhadap data *testing*.

- **Rule 4**

Pasien yang terdiagnosis stroke *haemorrhagic* tidak memiliki gejala yang tampak pada saat pasien tertidur (RSLEEP=N), tidak memiliki kelainan detak jantung (RATRIAL=N). Tidak melakukan proses CT-scan (RCT=N). Gejala yang tampak adalah defisit pada wajah (RDEF1=Y) dan defisit pada kaki (RDEF3=Y). Akurasi *rule* ini sebesar 9.884% terhadap data *testing*.

- **Rule 5**

Pasien yang terdiagnosis stroke *haemorrhagic* tidak memiliki gejala yang tampak pada saat pasien tertidur (RSLEEP=N), tidak memiliki kelainan detak jantung (RATRIAL=N). Tidak melakukan proses CT-scan (RCT=N) dan tidak terlihat infark pada bagian kepala (RVISINF=N). Gejala yang tampak adalah defisit pada lengan/tangan (RDEF2=Y) dan defisit pada kaki (RDEF3=Y). Akurasi *rule* ini sebesar 33.140% terhadap data *testing*.

- **Rule 6**

Pasien yang terdiagnosis stroke *haemorrhagic* tidak memiliki gejala yang tampak pada saat pasien tertidur (RSLEEP=N), tidak memiliki kelainan detak jantung (RATRIAL=N). Tidak melakukan proses CT-scan (RCT=N). Gejala yang tampak adalah defisit pada lengan/tangan (RDEF2=Y) dan defisit pada kaki (RDEF3=Y). Akurasi *rule* ini sebesar terhadap data *testing*.

Faktor – faktor yang berpengaruh terhadap diagnosis stroke *haemorrhagic* adalah sebagai berikut :

1. Pasien tidak memiliki gejala yang tampak pada saat tidur (RSLEEP=N). Faktor ini muncul sebanyak 6 kali / terdapat di dalam 6 *rule*.
2. Pasien tidak memiliki kelainan detak jantung (RATRIAL=N). Faktor ini muncul sebanyak 6 kali / terdapat di dalam 6 *rule*.
3. Pasien tidak melakukan proses CT-scan (RCT=N). Faktor ini muncul sebanyak 6 kali/terdapat di dalam 6 *rule*.
4. Pasien tidak memiliki infark yang terlihat (RVISINF=N). Faktor ini muncul sebanyak 3 kali / terdapat di dalam 3 *rule*.
5. Pasien memiliki gejala defisit pada wajah (RDEF1=Y). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.
6. Pasien memiliki gejala defisit pada tangan/lengan (RDEF2=Y). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.
7. Pasien memiliki gejala defisit pada kaki (RDEF3=Y). Faktor ini muncul sebanyak 6 kali / terdapat di dalam 6 *rule*.

c. *Indeterminate*

Pasien dalam kelompok ini belum bisa ditentukan jenis penyakit strokenya. Tabel 6-3 berisi detail dari *rule* yang terbentuk dan nilai akurasi dari masing – masing *rule* terhadap data *testing*.

Tabel 6-3 Rule Indeterminate

No	<i>RULE</i>	Akurasi
1	{RCT=N,RVISINF=N,RDEF1=Y,RDEF2=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	32.203%
2	{RCT=N,RDEF1=Y,RDEF2=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	32.203%
3	{RSLEEP=N,RATRIAL=N,RCT=N,RVISINF=N,RDEF1=Y,RDEF3=Y} => {DIAGNOSIS=Indeterminate}	25.424%
4	{SEX=F,RCT=N,RDEF1=Y} => {DIAGNOSIS=Indeterminate}	35.254%
5	{RCT=N,RVISINF=N,RDEF1=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	33.898%
6	{RCT=N,RDEF1=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	33.898%
7	{RCT=N,RVISINF=N,RDEF2=Y,RDEF3=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	33.898%

- **Rule 1**

Pasien yang terdiagnosis stroke *indeterminate* tidak melakukan CT-scan (RCT=N) dan tidak terlihat infark (RVISINF=N). Gejala yang diderita adalah adanya defisit pada wajah (RDEF1=Y), defisit pada tangan/lengan (RDEF2=Y) dan kelainan pemahanan bahasa / *dysphasia* (RDEF4=Y). Akurasi *rule* ini adalah sebesar 32.203% pada data *testing*.

- **Rule 2**

Pasien yang terdiagnosis stroke *indeterminate* tidak melakukan CT-scan ($RCT=N$). Gejala yang diderita adalah adanya defisit pada wajah ($RDEF1=Y$), defisit pada tangan/ lengan ($RDEF2=Y$) dan kelainan pemahanan bahasa / *dysphasia* ($RDEF4=Y$). Akurasi *rule* ini adalah sebesar 32.203% pada data *testing*.

- **Rule 3**

Pasien yang terdiagnosis stroke *indeterminate* tidak memiliki gejala yang tampak pada saat tidur ($RSLEEP=N$) dan tidak memiliki kelainan detak jantung ($RATRIAL=N$). Pasien tidak melakukan CT-scan ($RCT=N$) dan tidak terlihat adanya infark ($RVISINF=N$). Gejala yang diderita adalah adanya defisit pada wajah ($RDEF1=Y$) dan defisit pada kaki ($RDEF3=Y$). Akurasi *rule* ini adalah sebesar 25.424% pada data *testing*.

- **Rule 4**

Pasien yang terdiagnosis stroke *indeterminate* berjenis kelamin perempuan ($SEX=F$), tidak melakukan proses CT-scan ($RCT=N$) dan memiliki gejala defisit pada wajah ($RDEF1=Y$). Akurasi *rule* ini adalah sebesar 35.254% pada data *testing*.

- **Rule 5**

Pasien yang terdiagnosis stroke *indeterminate* tidak melakukan CT-scan ($RCT=N$) dan tidak terlihat adanya infark ($RVISINF=N$). Gejala yang diderita adalah defisit pada wajah ($RDEF1=Y$) dan kelainan pemahanan bahasa / *dysphasia* ($RDEF4=Y$). Akurasi *rule* ini adalah sebesar 33.898% pada data *testing*.

- **Rule 6**

Pasien yang terdiagnosis stroke *indeterminate* tidak melakukan CT-scan ($RCT=N$). Gejala yang diderita adalah defisit pada wajah ($RDEF1=Y$) dan kelainan

pemahaman bahasa / *dysphasia* (RDEF4=Y). Akurasi *rule* ini adalah sebesar 33.898% pada data *testing*.

• **Rule 7**

Pasien yang terdiagnosis stroke *indeterminate* tidak melakukan CT-scan (RCT=N) dan tidak terlihat adanya infark (RVISINF=N). Gejala yang tampak adalah adanya defisit pada tangan/lengan (RDEF2=Y), defisit pada kaki (RDEF3=Y) dan kelainan pemahaman bahasa / *dysphasia* (RDEF4=Y). Akurasi *rule* ini adalah sebesar 33.898% pada data *testing*.

Faktor – faktor yang berpengaruh terhadap diagnosis stroke *indeterminate* adalah sebagai berikut :

1. Pasien tidak melakukan CT-scan (RCT=N). Faktor ini muncul sebanyak 7 kali / terdapat di dalam 7 *rule*.
2. Tidak terlihat adanya infark (RVISINF=N). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.
3. Terdapat gejala defisit pada wajah (RDEF1=Y). Faktor ini muncul sebanyak 6 kali / terdapat di dalam 6 *rule*.
4. Terdapat gejala defisit pada tangan/lengan (RDEF2=Y). Faktor ini muncul sebanyak 3 kali / terdapat di dalam 3 *rule*.
5. Terdapat kelainan pemahaman bahasa / *dysphasia* (RDEF4=Y). Faktor ini muncul sebanyak 5 kali / terdapat di dalam 5 *rule*.
6. Tidak ada gejala yang Nampak pada saat pasien tidur (RSLEEP=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
7. Tidak adanya kelainan detak jantung (RATRIAL=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
8. Terdapat gejala defisit pada kaki (RDEF3=Y). Faktor ini muncul sebanyak 2 kali / terdapat di dalam 3 *rule*.

9. Jenis kelamin pasien adalah perempuan (SEX=F). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.

d. Not a stroke

Pasien digolongkan tidak menderita stroke apabila tidak menderita gejala – gejala stroke(RDEF6, RDEF4, RDEF5,RDEF), tidak ditemukan pendarahan otak (RVISINF) dan tidak ada kelainan detak jantung (RATRIAL). Tabel 6-4 berisi detail *rule* yang terbentuk dan nilai akurasi masing – masing *rule* terhadap data *testing*.

Tabel 6-4 Rule Not a stroke

No	RULE	Akurasi
1	{RATRIAL=N,RVISINF=N,RDEF6=N} => {DIAGNOSIS=Not a stroke}	52.555%
2	{RVISINF=N,RDEF4=N} => {DIAGNOSIS=Not a stroke}	54.015%
3	{RSLEEP=N,RATRIAL=N,RVISINF=N} => {DIAGNOSIS=Not a stroke}	56.934%
4	{RATRIAL=N,RVISINF=N,RDEF5=N} => {DIAGNOSIS=Not a stroke}	51.825%
5	{RVISINF=N,RDEF6=N} => {DIAGNOSIS=Not a stroke}	64.964%
6	{RCONSC=F,RATRIAL=N,RVISINF=N} => {DIAGNOSIS=Not a stroke}	62.774%
7	{RVISINF=N,RDEF5=N} => {DIAGNOSIS=Not a stroke}	64.234%

- **Rule 1**

Pada pasien yang tidak terdiagnosis stroke, pasien tidak memiliki kelainan detak jantung (RATRIAL=N), tidak memiliki infark (RVISINF=N) dan tidak mengalami kelainan visuospasial (RDEF6=N). Akurasi *rule ini* terhadap data *testing* adalah sebesar 52.56%.

- **Rule 2**

Pada pasien yang tidak terdiagnosis stroke tidak memiliki infark (RVISINF=N) dan tidak mengalami dysphasia (RDEF4=N). Akurasi *rule ini* terhadap data *testing* adalah sebesar 54.02%.

- **Rule 3**

Pada pasien yang tidak terdiagnosis stroke, pasien tidak memiliki gejala yang nampak pada saat tidur (RSLEEP=N), tidak ada kelainan detak jantung (RATRIAL=N) dan tidak memiliki infark (RVISINF=N). Akurasi *rule ini* terhadap data *testing* adalah sebesar 56.93%.

- **Rule 4**

Pada pasien yang tidak terdiagnosis stroke tidak ada kelainan detak jantung (RATRIAL=N), tidak memiliki infark (RVISINF=N) dan tidak ada gejala hemianopia / pengurangan kemampuan penglihatan (RDEF5=N). Akurasi *rule ini* terhadap data *testing* adalah sebesar 51.83%.

- **Rule 5**

Pada pasien yang tidak terdiagnosis stroke, pasien tidak memiliki infark (RVISINF=N) dan tidak ada kelainan visuospasial (RDEF6=N). Akurasi *rule ini* terhadap data *testing* adalah sebesar 64.96%.

- **Rule 6**

Pada pasien yang tidak terdiagnosis stroke, pasien dalam kondisi sangat sadar / *fully alert* (RCONSC=F), tidak ada kelainan detak jantung (RATRIAL=N) dan tidak

memiliki infark (RVISINF=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 62.77%.

- **Rule 7**

Pada pasien yang tidak terdiagnosis stroke, pasien tidak memiliki infark (RVISINF=N) dan tidak ada gejala hemianopia / pengurangan kemampuan penglihatan (RDEF5=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 64.23%.

Faktor – faktor yang mempengaruhi seseorang tidak terdiagnosis penyakit stroke / penyakit bukan stroke adalah sebagai berikut :

1. Pasien tidak memiliki kelainan detak jantung (RATRIAL=N). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.
2. Pasien tidak memiliki infark (RVISINF=N). Faktor ini muncul sebanyak 7 kali / terdapat di dalam 7 *rule*.
3. Pasien tidak mengalami gejala kelainan visuospasial (RDEF6=N). Faktor ini muncul sebanyak 2 kali / terdapat di dalam 2 *rule*.
4. Pasien tidak mengalami gejala dysphasia (RDEF4=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
5. Pasien tidak memiliki gejala yang tampak pada saat tidur (RSLEEP=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
6. Pasien tidak mengalami gejala hemianopia (RDEF5=N). Faktor ini muncul sebanyak 2 kali / terdapat di dalam 2 *rule*.
7. Pasien dalam keadaan sangat sadar / *fully alert* (RCONS=F). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.

e. Unknown

Pasien masuk dalam kategori ini jika tidak mengalami pendarahan otak (RVISINF) dan tidak terlihat gejala pada saat pasien tertidur (RSLEEP). Tabel 6-5 berisi detail *rule* dan nilai akurasi masing – masing *rule* terhadap data *testing*.

Tabel 6-5 Rule unknown stroke

No	RULE	Akurasi
1	{RVISINF=N} => {DIAGNOSIS=unknown}	60.000%
2	{RSLEEP=N} => {DIAGNOSIS=unknown}	70.000%
3	{RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=unknown}	90.000%

- **Rule 1**
Pasien tidak diketahui diagnosisnya jika tidak memiliki infark (RVISINF=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 60%.
- **Rule 2**
Pasien yang tidak diketahui diagnosisnya tidak memiliki gejala yang tampak pada saat pasien tidur (RSLEEP=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 70%.
- **Rule 3**
Pasien yang tidak diketahui diagnosisnya memiliki gejala defisit pada tangan/ lengan (RDEF2=Y) dan defisit pada bagian kaki (RDEF3=Y). Akurasi *rule* ini terhadap data *testing* adalah sebesar 90%. Pada kasus ini pasien biasanya menolak perawatan dan atau perekaman data lebih lanjut sehingga tidak terekam diagnosisnya.

Faktor – faktor yang mempengaruhi pasien pada kategori ini adalah sebagai berikut :

1. Pasien tidak memiliki infark (RVISINF=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
2. Pasien tidak memiliki gejala yang tampak pada saat tidur (RSLEEP=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
3. Pasien mengalami gejala defisit pada tangan/lengan (RDEF2=Y). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
4. Pasien mengalami gejala defisit pada kaki (RDEF3=Y). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.

6.2.2. Data Survival

Pada bagian ini akan dibahas *rule* yang muncul pada setiap kelas pada atribut OCCODE.

a. *Dead*

Pasien dalam uji coba ini dinyatakan meninggal. Di dalam tabel 6-6 terdapat detail *rule* dan nilai akurasi dari masing – masing *rule* terhadap data *testing*.

Tabel 6-6 Rule dead

No	<i>RULE</i>	Akurasi
1	{DAP=N,DOAC=N,DHAEMD=N} => {OCCODE=DEAD}	93.222%
2	{DAP=N,DOAC=N,DTHROMB=N} => {OCCODE=DEAD}	94.333%
3	{DAP=N,DOAC=N} => {OCCODE=DEAD}	96.111%
4	{DOAC=N,DTHROMB=N} => {OCCODE=DEAD}	96.667%
5	{DOAC=N} => {OCCODE=DEAD}	98.556%
6	{DAP=N,DHAEMD=N,DTHROMB=N} => {OCCODE=DEAD}	92.778%

No	RULE	Akurasi
7	{DAP=N,DHAEMD=N} => {OCCODE=DEAD}	94.667%
8	{DAP=N,DTHROMB=N} => {OCCODE=DEAD}	95.667%
9	{DAP=N} => {OCCODE=DEAD}	97.556%
10	{DTHROMB=N} => {OCCODE=DEAD}	98.000%

- **Rule 1**

Pasien yang meninggal tidak mengonsumsi obat antiplatelet percobaan (DAP=N), tidak mengonsumsi obat antikoagulan lain (DOAC=N) dan tidak mengalami kemaikan volume plasma darah / *haemodilution* (DHAEMD=N). Akurasi *rule* ini terhadap data testing adalah sebesar 93.22%.

- **Rule 2**

Pasien yang meninggal tidak mengonsumsi obat antiplatelet percobaan (DAP=N), tidak mengonsumsi obat antikoagulan lain (DOAC=N) dan tidak mengalami pemecahan gumpalan sumbatan darah / trombolisis (DTHROMB=N). Akurasi *rule* ini terhadap data testing adalah sebesar 94.33%.

- **Rule 3**

Pasien yang meninggal tidak mengonsumsi obat antiplatelet percobaan (DAP=N) dan tidak mengonsumsi obat antikoagulan lain (DOAC=N). Akurasi *rule* ini terhadap data testing adalah sebesar 96.11%.

- **Rule 4**

Pasien yang meninggal tidak mengonsumsi obat antikoagulan lain (DOAC=N) dan tidak mengalami pemecahan gumpalan sumbatan darah / trombolisis (DTHROMB=N). Akurasi *rule* ini terhadap data testing adalah sebesar 96.67%.

- **Rule 5**
Pasien yang meninggal tidak mengonsumsi obat antikoagulan lain (DOAC=N). Akurasi *rule* ini terhadap data testing adalah sebesar 98.56%.
- **Rule 6**
Pasien yang meninggal tidak mengonsumsi obat antiplatelet percobaan (DAP=N), tidak mengalami kemaikan volume plasma darah / *haemodilution* (DHAEMD=N) dan tidak mengalami pemecahan gumpalan sumbatan darah / trombolisis (DTHROMB=N). Akurasi *rule* ini terhadap data testing adalah sebesar 92.78%.
- **Rule 7**
Pasien yang meninggal tidak mengonsumsi obat antiplatelet percobaan (DAP=N) dan tidak mengalami kemaikan volume plasma darah / *haemodilution* (DHAEMD=N). Akurasi *rule* ini terhadap data testing adalah sebesar 94.67%.
- **Rule 8**
Pasien yang meninggal tidak mengonsumsi obat antiplatelet percobaan (DAP=N) dan tidak mengalami pemecahan gumpalan sumbatan darah / trombolisis (DTHROMB=N). Akurasi *rule* ini terhadap data testing adalah sebesar 95.67%.
- **Rule 9**
Pasien yang meninggal tidak mengonsumsi obat antiplatelet percobaan (DAP=N). Akurasi *rule* ini terhadap data testing adalah sebesar 97.56%.
- **Rule 10**
Pasien yang meninggal tidak mengalami pemecahan gumpalan sumbatan darah / trombolisis (DTHROMB=N). Akurasi *rule* ini terhadap data testing adalah sebesar 98%.

Faktor – faktor yang mempengaruhi kematian pasien adalah sebagai berikut:

1. Tidak ada konsumsi obat antiplatelet percobaan (DAP=N). Faktor ini muncul sebanyak 7 kali / terdapat di dalam 7 *rule*.
2. Tidak ada konsumsi obat ontikoagulan lain (DOAC=N). Faktor ini muncul sebanyak 5 kali / terdapat di dalam 5 *rule*.
3. Tidak adanya penambahan volume plasma darah (DHAEMD=N). Faktor ini muncul sebanyak 3 kali / terdapat di dalam 3 *rule*.
4. Tidak ada pemecahan gumpalan sumbatan darah / trombolisis (DTHROMB=N). Faktor ini muncul sebanyak 5 kali / terdapat di dalam 5 *rule*.

b. Recover

Pada bagian ini kita akan melihat *rule* dari kelas recover pada atribut OCCODE. Tabel 6-7 berisi detail *rule* dan nilai akurasi dari masing – masing *rule* terhadap data *testing*.

Tabel 6-7 Rule recover

No	Rule	Akurasi
1	{DSCH=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=RECOVER}	96.835%
2	{DSCH=N,DTHROMB=N,DCAUSE=unknown} => {OCCODE=RECOVER}	96.978%
3	{DMAJNCH=N,DCAUSE=unknown,CMPLHEP=Y} => {OCCODE=RECOVER}	96.691%
4	{DCAUSE=unknown,CMPLHEP=Y} => {OCCODE=RECOVER}	96.978%
5	{DSCH=N,DCAUSE=unknown} => {OCCODE=RECOVER}	97.122%

No	Rule	Akurasi
6	{DMAJNCH=N,DCAUSE=unknown,CMPLASP=Y} => {OCCODE=RECOVER}	97.122%
7	{DTHROMB=N,DMAJNCH=N,DCAUSE=unknown,CMPLHEP=Y} => {OCCODE=RECOVER}	95.396%
8	{DTHROMB=N,DCAUSE=unknown,CMPLHEP=Y} => {OCCODE=RECOVER}	95.683%
9	{DTHROMB=N,DCAUSE=unknown,CMPLASP=Y} => {OCCODE=RECOVER}	95.827%
10	{DSCH=N,DTHROMB=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=RECOVER}	96.691%

- **Rule 1**

Pasien yang sembuh dari stroke, pada masa perawatan tidak mengonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N), tidak mengalami pendarahan otak yang besar (DMAJNCH=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi dari *rule* ini terhadap data testing adalah sebesar 96.84%.

- **Rule 2**

Pasien yang sembuh dari stroke, pada masa perawatan tidak mengonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N), tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi dari *rule* ini terhadap data testing adalah sebesar 96.98%.

- **Rule 3**

Pasien yang sembuh dari stroke tidak mengalami pendarahan otak yang besar (DMAJNCH=N), sebab

kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown) dan mengonsumsi heparin sesuai aturan (komplit) (CMPLHEP=Y). Akurasi dari *rule* ini terhadap data testing adalah sebesar 96.69%.

- **Rule 4**

Pasien yang sembuh dari stroke sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown) dan mengonsumsi heparin sesuai aturan (komplit) (CMPLHEP=Y). Akurasi dari *rule* ini terhadap data testing adalah sebesar 96.98%

- **Rule 5**

Pasien yang sembuh dari stroke tidak mengonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi dari *rule* ini terhadap data testing adalah sebesar 97.12%.

- **Rule 6**

Pasien yang sembuh dari stroke tidak mengalami pendarahan otak yang besar (DMAJNCH=N), sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown) dan telah mengonsumsi aspirin sesuai dengan aturan (CMPLASP=Y). Akurasi dari *rule* ini terhadap data testing adalah sebesar 97.12%.

- **Rule 7**

Pasien yang sembuh dari stroke tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N), tidak mengalami pendarahan otak yang besar (DMAJNCH=N), sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown) dan mengonsumsi heparin sesuai aturan (komplit) (CMPLHEP=Y). Akurasi dari *rule* ini terhadap data testing adalah sebesar 95.40%.

- **Rule 8**

Pasien yang sembuh dari stroke tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N), sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown) dan mengkonsumsi heparin sesuai aturan (komplit) (CMPLHEP=Y). Akurasi dari *rule* ini terhadap data testing adalah sebesar 95.83%.

- **Rule 9**

Pasien yang sembuh dari stroke tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N), sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown) dan telah mengkonsumsi aspirin sesuai dengan aturan (CMPLASP=Y). Akurasi dari *rule* ini terhadap data testing adalah sebesar 95.83%.

- **Rule 10**

Pasien yang sembuh dari stroke tidak mengkonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N), tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N), tidak mengalami pendarahan otak yang besar (DMAJNCH=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi dari *rule* ini terhadap data testing adalah sebesar 96.69%.

Faktor – faktor yang berpengaruh terhadap kematian pasien stroke adalah sebagai berikut:

1. Pasien tidak mengkonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.

2. Pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N). Faktor ini muncul sebanyak 5 kali / terdapat di dalam 5 *rule*.
3. Penyebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Faktor ini muncul sebanyak 10 kali / terdapat di dalam 10 *rule*.
4. Pasien tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N). Faktor ini muncul sebanyak 5 kali / terdapat di dalam 5 *rule*.
5. Pasien mengkonsumsi heparin sesuai aturan (komplit) (CMPLHEP=Y). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.
6. Pasien mengkonsumsi aspirin sesuai dengan aturan (komplit) (CMPLASP=Y). Faktor ini muncul sebanyak 2 kali / terdapat di dalam 2 *rule*.

c. *Dependent*

Pasien dalam kelompok ini, pada akhir masa pemantauan 6 bulan memiliki kondisi masih bergantung kepada orang lain atau belum mandiri dikarenakan masih ada kecacatan. Detail *rule* dan nilai akurasi masing – masing *rule* terhadap data *testing* terdapat pada tabel 6-8.

Tabel 6-8 *Rule dependent*

No	<i>Rule</i>	Akurasi
1	{DIVH=N,DTHROMB=N} => {OCCODE=DEPENDENT}	96.216%
2	{DHAEMD=N,DMAJNCH=N} => {OCCODE=DEPENDENT}	96.396%

No	Rule	Akurasi
3	{DTHROMB=N,DMAJNCH=N} => {OCCODE=DEPENDENT}	97.057%
4	{DTHROMB=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=DEPENDENT}	94.955%

- **Rule 1**

Pasien pada kategori dependen tidak mengonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / *intravenous* (DIVH=N) dan tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 92.216%.

- **Rule 2**

Pasien pada kategori dependen tidak mengalami peningkatan volume plasma darah / *haemodilution* (DHAEMD=N) dan tidak mengalami pendarahan otak yang besar (DMAJNCH=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.40%.

- **Rule 3**

Pasien pada kategori dependen tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N) dan tidak mengalami pendarahan otak yang besar (DMAJNCH=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 97.06%.

- **Rule 4**

Pasien pada kategori dependen tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N), tidak mengalami pendarahan otak yang besar (DMAJNCH=N) dan sebab kematian yang tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi *rule* ini terhadap data *testing* adalah sebesar 94.96%

Faktor – faktor yang mempengaruhi pasien dalam kategori ini adalah sebagai berikut:

1. Pasien tidak mengkonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / *intravenous* (DIVH=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
2. Pasien tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N). Faktor ini muncul sebanyak 3 kali / terdapat di dalam 3 *rule*.
3. Pasien tidak mengalami peningkatan volume plasma darah / *haemodilution* (DHAEMD=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.
4. Pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N). Faktor ini muncul sebanyak 3 kali / terdapat di dalam 3 *rule*.
5. Sebab kematian pasien tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.

d. *Not recover*

Pasien dalam kelompok ini masih belum sembuh dari penyakit stroke. Faktor – faktor yang mempengaruhi kondisi pasien pada kelompok ini hampir sama dengan pasien pada kategori sebelumnya (*dependent*). Pada tabel 6-9 terdapat detail *rule* dan nilai akurasi dari masing – masing *rule* terhadap data *testing*.

Tabel 6-9 *Rule not recover*

No	<i>Rule</i>	Akurasi
1	{DMAJNCH=N,DCAUSE=unknown,CMPLASP=Y} => {OCCODE=NRECOVER}	95.734%

No	Rule	Akurasi
2	{DCAUSE=unknown,CMPLASP=Y} => {OCCODE=NRECOVER}	96.110%
3	{DSCH=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	96.487%
4	{DIVH=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	96.236%
5	{DSCH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	96.738%
6	{DMAJNCH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	99.624%
7	{DIVH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	96.487%
8	{DOAC=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	95.232%
9	{DCAUSE=unknown} => {OCCODE=NRECOVER}	100.000%
10	{DMAJNCH=N,CMPLASP=Y} => {OCCODE=NRECOVER}	95.734%
11	{CMPLASP=Y} => {OCCODE=NRECOVER}	96.110%
12	{DSCH=N,DMAJNCH=N} => {OCCODE=NRECOVER}	96.487%
13	{DIVH=N,DMAJNCH=N} => {OCCODE=NRECOVER}	96.236%
14	{DMAJNCH=N} => {OCCODE=NRECOVER}	99.624%
15	{DSCH=N} => {OCCODE=NRECOVER}	96.738%

No	Rule	Akurasi
16	{DIVH=N} => {OCCODE=NRECOVER}	96.487%

- **Rule 1**

Pada kategori ini pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N), sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown) dan telah mengkonsumsi aspirin sesuai dengan aturan (CMPLASP=Y). Akurasi *rule* ini terhadap data *testing* adalah sebesar 95.73%.

- **Rule 2**

Pada kategori ini pasien sebab kematiannya tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown) dan telah mengkonsumsi aspirin sesuai dengan aturan (CMPLASP=Y). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.11%.

- **Rule 3**

Pada kategori ini pasien tidak mengkonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N), tidak mengalami pendarahan otak yang besar (DMAJNCH=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.49%.

- **Rule 4**

Pada kategori ini pasien tidak mengkonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / *intravenous* (DIVH=N), tidak mengalami pendarahan otak yang besar (DMAJNCH=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.24%.

- **Rule 5**

Pada kategori ini pasien tidak mengonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.74%.

- **Rule 6**

Pada kategori ini pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi *rule* ini terhadap data *testing* adalah sebesar 99.62%.

- **Rule 7**

Pada kategori ini pasien tidak mengonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / *intravenous* (DIVH=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.49%.

- **Rule 8**

Pada kategori ini pasien tidak mengonsumsi obat antikoagulan lain (DOAC=N) dan sebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi *rule* ini terhadap data *testing* adalah sebesar 95.23%.

- **Rule 9**

Pada kategori ini pasien sebab kematiannya tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Akurasi *rule* ini terhadap data *testing* adalah sebesar 100%.

- **Rule 10**

Pada kategori ini pasien pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N) dan telah mengonsumsi aspirin sesuai dengan aturan (komplit)

(CMPLASP=Y). Akurasi *rule* ini terhadap data *testing* adalah sebesar 95.73%.

- **Rule 11**

Pada kategori ini pasien telah mengkonsumsi aspirin sesuai dengan aturan (kompli) (CMPLASP=Y). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.11% .

- **Rule 12**

Pada kategori ini pasien tidak mengkonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N) dan tidak mengalami pendarahan otak yang besar (DMAJNCH=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.49%.

- **Rule 13**

Pada kategori ini pasien tidak mengkonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / *intravenous* (DIVH=N) dan tidak mengalami pendarahan otak yang besar (DMAJNCH=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.24%.

- **Rule 14**

Pada kategori ini pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 99.62%.

- **Rule 15**

Pada kategori ini pasien tidak mengkonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.74%.

- **Rule 16**

Pada kategori ini pasien tidak mengkonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / *intravenous* (DIVH=N). Akurasi *rule* ini terhadap data *testing* adalah sebesar 96.49%.

Faktor – faktor yang mempengaruhi pasien pada kategori ini adalah sebagai berikut:

1. Tidak adanya pendarahan otak yang besar (DMAJNCH=N). Faktor ini muncul sebanyak 8 kali / terdapat di dalam 8 *rule*.
2. Sebab kematian pasien tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown). Faktor ini muncul sebanyak 9 kali / terdapat di dalam 9 *rule*.
3. Pasien telah mengkonsumsi aspirin sesuai dengan aturan (komplit) (CMPLASP=Y). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.
4. Pasien tidak mengkonsumsi heparin yang diberikan dibawah kulit / *underskin* (DSCH=N). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.
5. Pasien tidak mengkonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / *intravenous* (DIVH=N). Faktor ini muncul sebanyak 4 kali / terdapat di dalam 4 *rule*.
6. Pasien tidak mengkonsumsi obat antikoagulan lain (DOAC=N). Faktor ini muncul sebanyak 1 kali / terdapat di dalam 1 *rule*.

e. Missing status

Untuk pasien dengan status yang tidak diketahui atau missing, tidak bisa didapatkan *rule* atau polanya. Hal ini dapat disebabkan oleh data pasien dalam kategori ini yang jumlahnya sangat terbatas. Sehingga dengan minimum *support* / *confidence* yang kecil sekalipun sistem masih tidak dapat mengeluarkan *rule* atau polanya.

6.2.3. Pembobotan faktor – faktor yang berpengaruh terhadap data diagnosis dan data survival

Untuk mengetahui bobot dari faktor yang mempengaruhi dalam data diagnosis dan data survival. Akan dilakukan penghitungan bobot pada masing – masing faktor. Nilai bobot didapatkan dari rasio antara jumlah kemunculan faktor dan jumlah dari *rule* dalam kelas faktor tersebut.

Tabel dibawah ini akan berisi detail dari bobot masing – masing faktor beserta kelas yang dipengaruhi faktor tersebut. Tabel akan dibagi menjadi dua, yaitu table untuk data diagnosis dan table untuk data survival. Faktor yang memiliki bobot diatas 50% dianggap yang paling memiliki pengaruh terhadap kelas dari faktor yang bersangkutan.

Tabel 6-10 Faktor yang mempengaruhi diagnosis

Kelas (Jenis Stroke)	Faktor	Bobot%
<i>Ischaemic</i>	Tidak adanya kelainan detak jantung pada pasien (RATRIAL=N)	14%
<i>Ischaemic</i>	Pasien telah melakukan proses CT-scan (RCT=Y).	14%
<i>Ischaemic</i>	Pasien mengalami gejala defisit pada wajah (RDEF1=Y)	29%
<i>Ischaemic</i>	Pasien mengalami defisit pada kaki (RDEF3=Y)	29%
<i>Ischaemic</i>	Pasien tidak memiliki tanda pendarahan pada otak (RDEF7=N)	29%
<i>Ischaemic</i>	Pasien dalam keadaan sadar (RCONSC=Y)	14%
<i>Ischaemic</i>	Pasien dalam keadaan sangat sadar / <i>fully alert</i> (RCONSC=F)	29%
<i>Ischaemic</i>	Pasien tidak mengalami kelainan visuospasial (RDEF6=N)	29%

Kelas (Jenis Stroke)	Faktor	Bobot%
<i>Haemorrhagic</i>	Pasien tidak memiliki gejala yang tampak pada saat tidur (RSLEEP=N)	100%
<i>Haemorrhagic</i>	Pasien tidak memiliki kelainan detak jantung (RATRIAL=N)	100%
<i>Haemorrhagic</i>	Pasien tidak melakukan proses CT-scan (RCT=N)	100%
<i>Haemorrhagic</i>	Pasien tidak memiliki infark yang terlihat (RVISINF=N)	50%
<i>Haemorrhagic</i>	Pasien memiliki gejala defisit pada wajah (RDEF1=Y)	67%
<i>Haemorrhagic</i>	Pasien memiliki gejala defisit pada tangan/lengan (RDEF2=Y)	67%
<i>Haemorrhagic</i>	Pasien memiliki gejala defisit pada kaki (RDEF3=Y)	100%
<i>Indeterminate</i>	Pasien tidak melakukan CT-scan (RCT=N)	100%
<i>Indeterminate</i>	Tidak terlihat adanya infark (RVISINF=N)	57%
<i>Indeterminate</i>	Terdapat gejala defisit pada wajah (RDEF1=Y)	86%
<i>Indeterminate</i>	Terdapat gejala defisit pada tangan/lengan (RDEF2=Y)	43%
<i>Indeterminate</i>	Terdapat kelainan pemahanan bahasa / dysphasia (RDEF4=Y)	71%
<i>Indeterminate</i>	Tidak ada gejala yang Nampak pada saat pasien tidur (RSLEEP=N)	14%
<i>Indeterminate</i>	Tidak adanya kelainan detak jantung (RATRIAL=N)	14%
<i>Indeterminate</i>	Terdapat gejala defisit pada kaki (RDEF3=Y)	43%
<i>Indeterminate</i>	Jenis kelamin pasien adalah perempuan (SEX=F)	14%
<i>Not a stroke</i>	Pasien tidak memiliki kelainan detak jantung (RATRIAL=N)	57%

Kelas (Jenis Stroke)	Faktor	Bobot%
<i>Not a stroke</i>	Pasien tidak memiliki infark (RVISINF=N)	100%
<i>Not a stroke</i>	Pasien tidak mengalami gejala kelainan visuospasial (RDEF6=N)	29%
<i>Not a stroke</i>	Pasien tidak mengalami gejala dysphasia (RDEF4=N)	14%
<i>Not a stroke</i>	Pasien tidak memiliki gejala yang tampak pada saat tidur (RSLEEP=N)	14%
<i>Not a stroke</i>	Pasien tidak mengalami gejala hemianopia (RDEF5=N).	29%
<i>Not a stroke</i>	Pasien dalam keadaan sangat sadar / fully alert (RCONS=F)	14%
<i>Unknown</i>	Pasien tidak memiliki infark (RVISINF=N)	33%
<i>Unknown</i>	Pasien tidak memiliki gejala yang tampak pada saat tidur (RSLEEP=N)	33%
<i>Unknown</i>	Pasien mengalami gejala defisit pada tangan/lengan (RDEF2=Y).	33%
<i>Unknown</i>	Pasien mengalami gejala defisit pada kaki (RDEF3=Y)	33%

Tabel 6-11 Faktor yang mempengaruhi kondisi akhir pasien

Kelas (Kondisi akhir pasien)	Faktor	Bobot %
<i>Dead</i>	Tidak ada konsumsi obat antiplatelet percobaan (DAP=N)	70%
<i>Dead</i>	Tidak ada konsumsi obat ontikoagulan lain (DOAC=N)	50%
<i>Dead</i>	Tidak adanya penambahan volume plasma darah (DHAEMD=N)	30%
<i>Dead</i>	Tidak ada pemecahan gumpalan sumbatan darah / trombolisis (DTHROMB=N)	50%
<i>Recover</i>	Pasien tidak mengkonsumsi heparin yang diberikan dibawah kulit / underskin (DSCH=N)	40%
<i>Recover</i>	Pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N).	50%
<i>Recover</i>	Penyebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown).	100%
<i>Recover</i>	Pasien tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N)	50%
<i>Recover</i>	Pasien mengkonsumsi heparin sesuai aturan (kompli) (CMPLHEP=Y)	40%
<i>Recover</i>	Pasien mengkonsumsi aspirin sesuai dengan aturan (kompli) (CMPLASP=Y).	20%
<i>Dependent</i>	Pasien tidak mengkonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / intravenous (DIVH=N)	25%
<i>Dependent</i>	Pasien tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N)	75%

Kelas (Kondisi akhir pasien)	Faktor	Bobot %
<i>Dependent</i>	Pasien tidak mengalami peningkatan volume plasma darah / haemodilution (DHAEMD=N)	25%
<i>Dependent</i>	Pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N)	75%
<i>Dependent</i>	Sebab kematian pasien tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown)	25%
<i>Not Recover</i>	Tidak adanya pendarahan otak yang besar (DMAJNCH=N)	50%
<i>Not Recover</i>	Sebab kematian pasien tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown)	56%
<i>Not Recover</i>	Pasien telah mengkonsumsi aspirin sesuai dengan aturan (komplit) (CMPLASP=Y)	25%
<i>Not Recover</i>	Pasien tidak mengkonsumsi heparin yang diberikan dibawah kulit / underskin (DSCH=N)	25%
<i>Not Recover</i>	Tidak mengkonsumsi heparin percobaan yang disuntikkan kedalam pembuluh darah / intravenous (DIVH=N)	25%
<i>Not Recover</i>	Pasien tidak mengkonsumsi obat antikoagulan lain (DOAC=N)	6%

BAB VII

KESIMPULAN DAN SARAN

Pada bagian ini akan dijelaskan kesimpulan dari penelitian tugas akhir dan saran yang didapat berdasarkan pengerjaan tugas akhir.

7.1. Kesimpulan

Kesimpulan yang didapatkan dalam penelitian ini akan dibagi menjadi dua bagian, yang pertama adalah kesimpulan dari proses pengerjaan *association rule mining* dan yang kedua adalah kesimpulan mengenai hasil yang berupa list faktor – faktor yang berpengaruh pada data diagnosis dan data survival.

7.1.1. Proses *association rule mining*

1. Data dibagi menjadi dua, yaitu data diagnosis dan survival. Setiap data dibagi menjadi data *training* dan data *testing*, masing – masing sebesar 70% dan 30%.
2. Distribusi kelas dalam atribut dapat mempengaruhi kualitas data. Data yang memiliki distribusi kelas yang merata memiliki kualitas lebih baik. Dibuktikan dengan rata – rata nilai akurasi data survival lebih baik daripada data diagnosis.
3. Penentuan minimum *support* dan *confidence* mempengaruhi banyaknya rule yang dapat dihasilkan,. Minimum *support* dan minimum *confidence* ditentukan masing – masing sebesar 50% dan 60% dari jumlah transaksi masing – masing kelas.
4. Penggalan rule pada data diagnosis perlu dilakukan secara terpisah karena atribut yang ada di dalam data tersebut hanya memiliki korelasi kuat terhadap atribut hasil “Diagnosis” dan tidak berkorelasi dengan atribut hasil “OCCODE” pada data survival.

5. Selain itu atribut hasil “Diagnosis” merupakan input dari atribut hasil “OCCODE” yang menjelaskan tentang kondisi akhir pasien, sehingga jika diketahui kondisi akhir pasien dapat dilacak kembali diagnosis jenis stroke-nya (jenis stroke tercantum dalam atribut hasil “Diagnosis”) dan faktor – faktor yang mempengaruhi diagnosis jenis stroke tersebut juga dapat dilacak kembali dengan mudah .

7.1.2. Faktor – faktor yang berpengaruh pada data diagnosis dan data survival

1. Faktor – faktor yang berpengaruh pada data diagnosis dengan atribut hasil “Diagnosis” adalah sebagai berikut:
 - a. Stroke *Ischaemic*
 - Pasien mengalami gejala defisit pada wajah (RDEF1=Y)
 - Pasien mengalami defisit pada kaki (RDEF3=Y)
 - Pasien tidak memiliki tanda pendarahan pada otak (RDEF7=N)
 - Pasien dalam keadaan sangat sadar / fully alert (RCONSC=F)
 - Pasien tidak mengalami kelainan visuospasial (RDEF6=N)
 - b. Stroke *Haemorrhagic*
 - Pasien tidak memiliki gejala yang tampak pada saat tidur (RSLEEP=N)
 - Pasien tidak memiliki kelainan detak jantung (RATRIAL=N)
 - Pasien tidak melakukan proses CT-scan (RCT=N)
 - Pasien tidak memiliki infark yang terlihat (RVISINF=N)

- Pasien memiliki gejala defisit pada wajah (RDEF1=Y)
- Pasien memiliki gejala defisit pada tangan/lengan (RDEF2=Y)
- Pasien memiliki gejala defisit pada kaki (RDEF3=Y)

c. *Indeterminate*

- Pasien tidak melakukan CT-scan (RCT=N)
- Tidak terlihat adanya infark (RVISINF=N)
- Terdapat gejala defisit pada wajah (RDEF1=Y)
- Terdapat kelainan pemakanan bahasa / dysphasia (RDEF4=Y)

d. *Not a stroke* (bukan stroke)

Pasien terdiagnosis tidak menderita stroke apabila pasien tidak memiliki infark (RVISINF=N)

e. *unknown*

- Pasien tidak memiliki infark (RVISINF=N)
- Pasien tidak memiliki gejala yang tampak pada saat tidur (RSLEEP=N)
- Pasien mengalami gejala defisit pada tangan/lengan (RDEF2=Y).
- Pasien mengalami gejala defisit pada kaki (RDEF3=Y)

2. Faktor – faktor yang berpengaruh pada data survival dengan atribut hasil “OCCODE” adalah sebagai berikut:

a. *Dead* (meninggal)

- Tidak ada konsumsi obat antiplatelet percobaan (DAP=N)
- Tidak ada konsumsi obat antikoagulan lain (DOAC=N)
- Tidak ada pemecahan gumpalan sumbatan darah / trombolisis (DTHROMB=N)

- b. Recover* (sembuh)
- Pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N).
 - Penyebab kematian tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown).
 - Pasien tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N)
 - Pasien mengkonsumsi heparin sesuai aturan (komplit) (CMPLHEP=Y)
 - Pasien mengkonsumsi aspirin sesuai dengan aturan (komplit) (CMPLASP=Y).
- c. Dependent* (sembuh dan cacat)
- Pasien tidak mengalami pemecahan sumbatan darah / trombolisis (DTHROMB=N)
 - Pasien tidak mengalami peningkatan volume plasma darah / haemodilution (DHAEMD=N)
 - Pasien tidak mengalami pendarahan otak yang besar (DMAJNCH=N)
- d. Not recover* (belum sembuh)
- Tidak adanya pendarahan otak yang besar (DMAJNCH=N)
 - Sebab kematian pasien tidak diketahui karena pasien tidak meninggal (DCAUSE=unknown)
3. Akurasi rule pada data diagnosis memiliki rata – rata sebesar 49.30% dan data survival sebesar 96.44%. Hal ini disebabkan oleh distribusi kelas yang ada di dalam data survival yang lebih merata.

7.2. Saran

Berdasarkan hasil pengerjaan tugas akhir dan kesimpulan yang dapat diambil dari hasil, untuk meningkatkan dan mengembangkan bahasan pada tugas akhir ini. Penulis memberikan saran sebagai berikut:

1. Untuk hasil yang lebih baik, penelitian sebaiknya dilakukan dengan menggunakan data dengan proporsi kelas yang merata.
2. Algoritma apriori memang efektif untuk menggali *rule* dari data secara menyeluruh, namun untuk data berukuran besar prosesnya membutuhkan sumber daya yang lebih besar. Jadi diharapkan untuk menggunakan sumber daya yang mencukupi, seperti laptop yang memiliki spesifikasi baik.

DAFTAR PUSTAKA

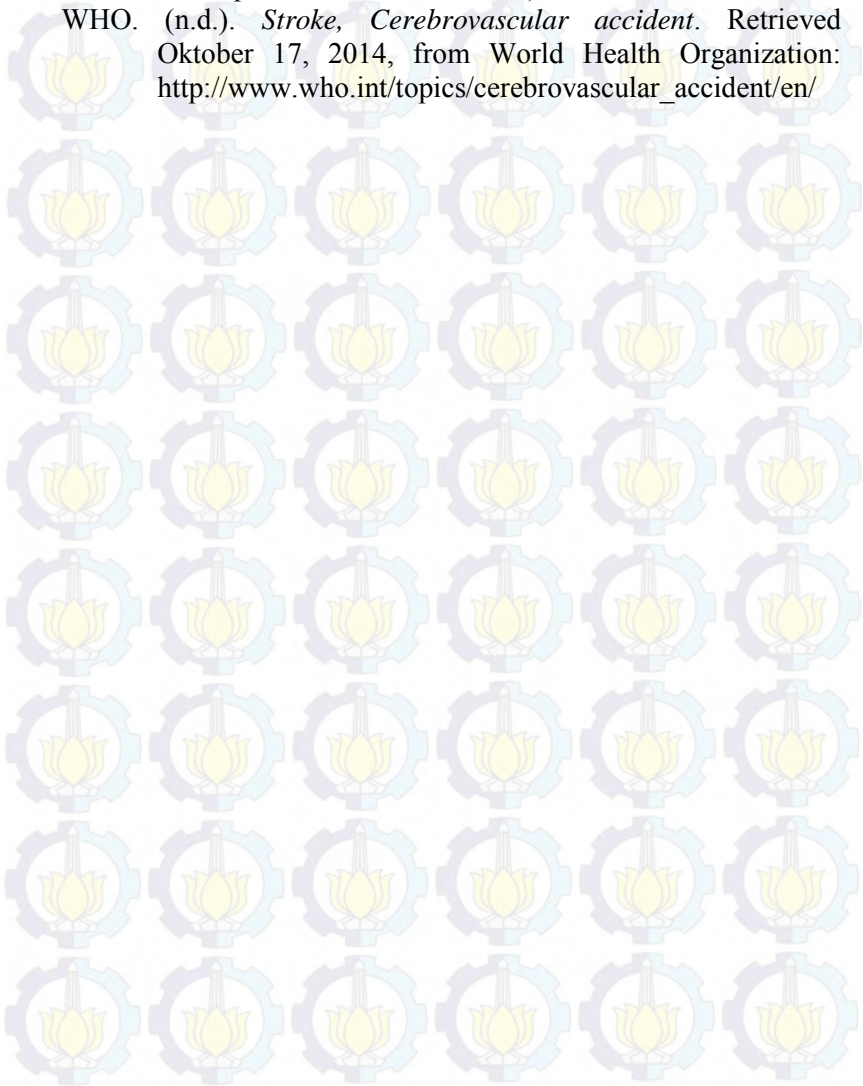
- Association rules*. (2014, December 2). Retrieved from RDataMining.com: R and Data Mining: <http://www.rdatamining.com/examples/association-rules>
- Badan Penelitian dan Pengembangan Kesehatan. (2013). *Riset Kesehatan Dasar Tahun 2013*. Jakarta: Badan Penelitian dan Pengembangan Kesehatan Kementerian Kesehatan RI.
- Budi. (2010, April 17). *Data Mining: Market Basket Analysis for Marketing Strategy*. Retrieved 10 30, 2014, from Statistika Komputasi: <http://statistikakomputasi.wordpress.com/2010/04/17/data-mining-market-basket-analysis-for-marketing-strategy/>
- CVA. (2014, 11 31). Retrieved from MedRevise: <http://www.medrevise.co.uk/index.php?title=CVA>
- Dewi, d. (n.d.). *Jenis - jenis stroke*. Retrieved Oktober 17, 2014, from Poliklinik IPDN: <http://poliklinik.ipdn.ac.id/home/artikel-kesehatan/jenis---jenis-stroke>
- Hahsler, M. (2014, 11 11). *A Comparison of Commonly Used Interest Measures for Association Rules*. Retrieved from Michael Hahsler: http://michael.hahsler.net/research/association_rules/measures.html
- Ishibuchi, H., Kuwajima, I., & Nojima, Y. (2007). Prescreening of candidate rules using association rule mining and pareto optimality in genetic rule selection. *Knowledge-based intelligent information and engineering systems*, 509-516.
- Mackay, J., & Mensah, A. G. (n.d.). *The Atlas of Heart Disease and Stroke*. Retrieved Oktober 19, 2014, from World Health Organization: http://www.who.int/cardiovascular_diseases/en/cvd_atlas_15_burden_stroke.pdf?ua=1
- Minimum Apriori*. (2013, Mei 16). Retrieved Oktober 17, 2014, from Nurfadillah100:

<http://nurfadillah100.blogspot.com/2013/05/minimum-apriori.html>

- Nahar, J., Imam, T., Tickle, K. S., & Chen, Y.-p. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, 1086 - 1093.
- Okatoria, R., Maharani, W., & Firdaus, Y. (2010). Content Based Recommender System Menggunakan Algoritma Apriori. *Konferensi Nasional Sistem dan Informatika*.
- Rambe, S. A. (2006, Desember 20). *STROKE: SEKILAS TENTANG DEFINISI, PENYEBAB, EFEK, DAN FAKTOR RESIKO*. Retrieved Oktober 16, 2014, from Departemen Neurologi FK-USU: [http://repository.usu.ac.id/bitstream/123456789/18925/1/i-km-des2006-10%20\(3\).pdf](http://repository.usu.ac.id/bitstream/123456789/18925/1/i-km-des2006-10%20(3).pdf)
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In *Encyclopedia of database systems* (pp. 532-538). USA: Springer.
- Rijsbergen, C. (1979). *Information Retrieval Glasgow*. Information retrieval group, University of Glasgow .
- Sandercock, P. A., Niewada, M., & Czlonkowska, A. (2011). The International Stroke Trial database. *trialsjournal*.
- Tan, P. N., Kumar, V., & Srivastava, J. (2004). Selecting the Right Objective Measure for Association Analysis. *Information System*, 293-313.
- Tan, P. N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. USA: Addison-Wesley.
- Tan, S. &. (2006). *Introduction to Data Mining*. Pearson Addison Wesley.
- University of Waikato. (n.d.). *Weka 3: Data Mining Software in Java*. Retrieved from www.cs.waikato.ac.nz:www.cs.waikato.ac.nz/ml/weka/
- Wandi, N., Hendrawan, R. A., & Mukhlason, A. (2012). Pengembangan Sistem Rekomendasi Penelusuran Buku dengan Penggalan Association Rule Menggunakan

Algoritma Apriori (Studi Kasus Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur). *Jurnal Teknik ITS*.

WHO. (n.d.). *Stroke, Cerebrovascular accident*. Retrieved Oktober 17, 2014, from World Health Organization: http://www.who.int/topics/cerebrovascular_accident/en/



LAMPIRAN

A.1. Contoh data dari dataset International Stroke Trial (IST)

RDE LAY	RCO NSC	SE X	A G E	RSL EEP	RATR IAL	R C T	RVIS INF	RD EF1	RD EF2	RD EF3	RD EF4	RD EF5	RD EF6	RD EF7	RD EF8	STY PE
A	D	M	B	Y		Y	Y	N	N	N	Y	N	Y	N	Y	PA CS
A	F	M	B	Y		Y	N	Y	Y	Y	N	N	N	N	N	LA CS
B	F	F	B	N		Y	N	Y	Y	Y	N	N	N	N	N	LA CS
A	F	M	A	N		N	N	N	N	N	Y	N	N	N	N	PA CS
A	F	M	A	N		N	N	Y	Y	Y	N	N	N	N	N	LA CS
B	F	M	B	N		Y	N	Y	Y	Y	N	N	N	N	N	LA CS
A	F	F	B	N		N	N	Y	Y	Y	N	N	N	Y	N	PO CS
A	F	M	B	N		Y	N	N	N	N	N	N	N	N	Y	OT H
A	F	M	B	N		N	N	N	N	N	N	N	N	N	Y	OT H
A	F	M	A	Y		N	N	N	N	N	N	N	Y	N	Y	PA CS

RDE LAY	RCO NSC	SE X	A G E	RSL EEP	RATR IAL	R C T	RVIS INF	RD EF1	RD EF2	RD EF3	RD EF4	RD EF5	RD EF6	RD EF7	RD EF8	STY PE
A	D	M	B	Y		Y	N	Y	Y	Y	N	N	N	Y	N	PO CS
A	F	F	B	N		Y	Y	Y	Y	Y	Y	N	N	N	N	PA CS
A	D	F	A	N		N	N	Y	Y	Y	Y	Y	N	N	N	TA CS
A	F	F	A	Y		Y	N	N	N	N	N	N	N	N	Y	OT H
A	F	M	B	N		N	N	N	N	N	N	N	N	N	Y	OT H
A	F	F	B	N		N	N	Y	Y	Y	Y	Y	C	N	N	TA CS
A	D	F	B	Y		Y	Y	Y	Y	Y	N	Y	Y	N	N	TA CS
B	F	M	B	N		Y	Y	Y	Y	Y	N	N	N	Y	N	PO CS
A	D	M	B	N		Y	Y	Y	Y	Y	N	Y	Y	N	Y	TA CS
A	F	M	B	N		N	N	Y	Y	Y	N	N	Y	N	N	PA CS
A	F	F	B	N		Y	Y	Y	Y	Y	N	Y	Y	N	N	TA CS
A	F	M	A	Y		N	N	Y	Y	Y	N	Y	Y	N	N	TA CS
A	D	M	A	N		N	N	Y	Y	Y	N	Y	Y	N	N	TA CS

RDE LAY	RCO NSC	SE X	A G E	RSL EEP	RATR IAL	R C T	RVIS INF	RD EF1	RD EF2	RD EF3	RD EF4	RD EF5	RD EF6	RD EF7	RD EF8	STY PE
A	F	M	B	Y		Y	N	N	N	N	N	N	N	N	Y	OT H

DAYL OCAL	RX HE P	DA SP1 4	DA SPL T	DL H1 4	DM H1 4	DH H1 4	OND RUG	DS C H	DI V H	D A P	D O AC	DG OR M	DS TE R	DC A A	DHA EMD	DCA REN D	DTH ROM B
Wedn esday	N	Y	Y	N		N	14D			N	N	N	N	N	N		
Satur ay	L	N	Y	Y		N	14D			N	N	N	N	Y	N		
Tuesd ay	N	Y	Y	N		N	11D			N	N	N	N	N	N		
Satur ay	H	N	Y	N		Y	14D			N	N	N	N	Y	N		
Friday	H	Y	Y	N		Y	14D			N	N	N	N	N	N		
Mond ay	N	N	Y	N		N	7D			N	N	N	N	N	N		
Sunda y	N	Y	Y	N		N	14D			N	N	N	N	N	N		
Mond ay	L	N	N	Y		N	11D			N	N	N	N	N	N		
Mond ay	H	Y	Y	N		Y	5D			N	N	N	N	N	N		
Mond ay	N	N	N	N		N	14D			N	N	N	N	N	N		
Satur ay	L	Y	Y	Y		N	11D			N	N	N	N	N	N		
Thurs day	N	N	Y	N		N	1D			N	N	N	N	N	N		
Friday	H	N	N	N		N	14D			N	N	N	N	N	N		
Satur ay	N	Y	Y	N		N	14D			N	N	N	N	N	N		

DAYL OCAL	RX HE P	DA SP1 4	DA SPL T	DL H1 4	DM H1 4	DH H1 4	OND RUG	DS C H	DI V H	D A P	D O AC	DG OR M	DS TE R	DC A A	DHA EMD	DCA REN D	DTH ROM B
Friday	L	Y	N	Y		N	6D			N	N	N	N	N	N		
Friday	N	N	Y	N		N	11D			N	N	N	N	N	N		
Mond ay	H	Y	Y	N		Y	14D			N	N	N	N	N	N		
Sunda y	H	N	Y	N		Y	9D			N	N	N	N	N	N		
Tuesd ay	N	Y	Y	N		N	14D			N	N	N	N	Y	N		
Tuesd ay	N	Y	N	N		N	10D			N	N	N	N	N	N		
Wedn esday	L	Y	Y	Y		N	7D			N	N	N	N	Y	N		
Sunda y	L	N	N	N		N	14D			N	Y	N	N	N	N		
Friday	L	N	Y	Y		N	14D			N	N	N	N	N	N		
Saturd ay	N	N	N	N		N	14D			N	N	N	N	N	N		

DMA JNCH	DMAJ NCHX	DS ID E	DR SIS C	DR SH	DRS UNK	D P E	DA LIV E	DPL ACE	DDEADC	FDE ADC	FPL ACE	F A P	FO AC	FU1 REC D	FU2_ DON E
N		N	N	N		N	N		unknown	unkn own	E			14	187
N		N	N	N		N	N		unknown	unkn own	A			16	192
N		N	N	N		N	Y		unknown	unkn own	A			11	189
N		N	N	N		N	Y		unknown	unkn own	A			23	183
N		N	N	N		N	N		unknown	unkn own	E			17	214
N		N	N	N		N	Y		unknown	unkn own	A			16	185
N		Y	N	N		N	N		unknown	unkn own	A			15	194
N		N	N	N		N	Y		unknown	unkn own	A			37	193
N		N	N	N		N	Y		unknown	unkn own	A			23	193
N		N	N	N		N	Y		unknown	unkn own	A			42	189
N		N	N	N		N	Y		unknown	unkn own	A			24	194
N		N	N	N		N	Y		unknown	unkn own	A			13	221
N		N	N	N		N	N		unknown	unkn own	D			24	242

DMA JNCH	DMAJ NCHX	DS ID E	DR SIS C	DR SH	DRS UNK	D P E	DA LIV E	DPL ACE	DDEADC	FDE ADC	FPL ACE	F A P	FO AC	FU1 REC D	FU2_ DON E
N		N	N	N		N	N		unknown	unkn own	U	U	U	38	614
N		N	N	N		N	Y		unknown	unkn own	U	U	U	17	608
N		N	N	N		N	Y		unknown	unkn own	A			17	206
N		N	N	N		N	N		unknown	unkn own	A			21	198
N		N	N	N		N	Y		unknown	unkn own	A			15	198
N		N	N	N		N	N		Pneumonia	Pneumonia				21	21
N		N	N	N		N	N		Other vascular/unk nown	Other vascular/unknown				36	687
N		N	N	N		N	Y		unknown	unkn own	A			20	193
N		N	N	N		N	N		unknown	unkn own	A			24	189
N		N	N	N		N	N		unknown	unkn own	E			26	215
N		N	N	N		N	N		unknown	unkn own	A			25	210

DIAGNOSIS	COUNTRY	NCCODE	CMPLASP	CMPLHEP	SET14D	ID14	ISC14	STRK14	HTI14	PE14	OCCODE
Ischaemic	UK	13	Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	RECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Not a stroke	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	RECOVER
Haemorrhagic	UK	4	Y	N	Y	N	N	N	N	N	NRECOVER

DIAGNOSIS	COUNTRY	NCCODE	CMPLASP	CMPLHEP	SET14D	ID14	ISC14	STRK14	HTI14	PE14	OCCODE
Not a stroke	UK		Y	Y	Y	N	N	N	N	N	missing status
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	missing status
Indeterminate	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	DEAD
Indeterminate	UK		Y	Y	Y	Y	N	N	N	N	DEAD
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK	11	Y	N	Y	N	N	N	N	N	DEPENDENT
Indeterminate	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Indeterminate	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	NRECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	NRECOVER

DIAGNOSIS	COUNTRY	NCCODE	CMPLAS	CMPLH	SET14	ID1	ISC1	STRK1	HTI1	PE1	OCCODE
Ischaemic	UK	E	Y	Y	Y	N	N	N	N	N	NRECOVER
Ischaemic	UK		Y	Y	Y	N	N	N	N	N	DEPENDENT

A.2. Hasil rule data diagnosis

No	RULE	Support	Confidence	Lift	Coverage	Akurasi
1	{RATRIAL=N,RCT=Y} => {DIAGNOSIS=Ischaemic}	0.515	0.973	1.086	0.530	58.175%
2	{RDEF1=Y,RDEF2=Y,RDEF7=N} => {DIAGNOSIS=Ischaemic}	0.516	0.906	1.012	0.570	56.891%
3	{RCONSC=Y,RDEF2=Y,RDEF7=N}=> {DIAGNOSIS=Ischaemic}	0.506	0.907	1.011	0.570	57.658%
4	{RCONSC=F,RDEF6=N} => {DIAGNOSIS=Ischaemic}	0.536	0.902	1.007	0.594	60.974%
5	{RDEF1=Y,RDEF3=Y} => {DIAGNOSIS=Ischaemic}	0.544	0.900	1.005	0.604	60.073%
6	{RCONSC=F} => {DIAGNOSIS=Ischaemic}	0.688	0.900	1.004	0.765	78.072%
7	{RDEF6=N} => {DIAGNOSIS=Ischaemic}	0.593	0.899	1.004	0.659	66.763%
8	{RSLEEP=N,RATRIAL=N,RCT=N,RVISINF=N,RDEF1=Y,RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	0.012	0.123	4.016	0.101	33.140%
9	{RSLEEP=N,RATRIAL=N,RCT=N,RDEF1=Y,RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	0.012	0.123	4.010	0.101	33.140%
10	{RSLEEP=N,RATRIAL=N,RCT=N,RVISINF=N,RDEF1=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	0.012	0.122	3.993	0.101	33.140%
11	{RSLEEP=N,RATRIAL=N,RCT=N,RDEF1=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	0.012	0.122	3.987	0.101	9.884%

No	RULE	Support	Confidence	Lift	Coverage	Akurasi
12	{RSLEEP=N,RATRIAL=N,RCT=N,RVISINF=N,RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	0.016	0.120	3.937	0.132	33.140%
13	{RSLEEP=N,RATRIAL=N,RCT=N,RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=Haemorrhagic}	0.016	0.120	3.933	0.132	44.186%
14	{RCT=N,RVISINF=N,RDEF1=Y,RDEF2=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	0.017	0.160	3.212	0.104	32.203%
15	{RCT=N,RDEF1=Y,RDEF2=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	0.017	0.160	3.207	0.104	32.203%
16	{RSLEEP=N,RATRIAL=N,RCT=N,RVISINF=N,RDEF1=Y,RDEF3=Y} => {DIAGNOSIS=Indeterminate}	0.017	0.160	3.227	0.132	25.424%
17	{SEX=F,RCT=N,RDEF1=Y} => {DIAGNOSIS=Indeterminate}	0.018	0.158	3.178	0.115	35.254%
18	{RCT=N,RVISINF=N,RDEF1=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	0.018	0.158	3.163	0.114	33.898%
19	{RCT=N,RDEF1=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	0.018	0.157	3.159	0.114	33.898%
20	{RCT=N,RVISINF=N,RDEF2=Y,RDEF3=Y,RDEF4=Y} => {DIAGNOSIS=Indeterminate}	0.017	0.155	3.108	0.108	33.898%
21	{RATRIAL=N,RVISINF=N,RDEF6=N} => {DIAGNOSIS=Not a stroke}	0.013	0.033	1.478	0.383	52.555%
22	{RVISINF=N,RDEF4=N} => {DIAGNOSIS=Not a stroke}	0.012	0.033	1.466	0.368	54.015%
23	{RSLEEP=N,RATRIAL=N,RVISINF=N} => {DIAGNOSIS=Not a stroke}	0.012	0.033	1.458	0.375	56.934%

No	<i>RULE</i>	<i>Support</i>	<i>Confidence</i>	<i>Lift</i>	<i>Coverage</i>	Akurasi
24	{RATRIAL=N,RVISINF=N,RDEF5=N} => {DIAGNOSIS=Not a stroke}	0.012	0.032	1.444	0.378	51.825%
25	{RVISINF=N,RDEF6=N} => {DIAGNOSIS=Not a stroke}	0.015	0.032	1.441	0.466	64.964%
26	{RCONSC=F,RATRIAL=N,RVISINF=N} => {DIAGNOSIS=Not a stroke}	0.014	0.032	1.439	0.439	62.774%
27	{RVISINF=N,RDEF5=N} => {DIAGNOSIS=Not a stroke}	0.014	0.031	1.401	0.455	64.234%
28	{RVISINF=N} => {DIAGNOSIS=unknown}	0.001	0.002	1.164	0.668	60.000%
29	{RSLEEP=N} => {DIAGNOSIS=unknown}	0.001	0.001	1.091	0.713	70.000%
30	{RDEF2=Y,RDEF3=Y} => {DIAGNOSIS=unknown}	0.001	0.001	1.054	0.738	90.000%
Rata - rata		0.140	0.282	2.223	0.358	49.310%

A.3. Hasil rule data survival

No	RULE	Support	Confidence	Lift	Coverage	Akurasi
1	{DAP=N,DOAC=N,DHAEMD=N} => {OCCODE=DEAD}	0.20	0.22	1.03	0.90	93.22%
2	{DAP=N,DOAC=N,DTHROMB=N} => {OCCODE=DEAD}	0.20	0.22	1.03	0.91	94.33%
3	{DAP=N,DOAC=N} => {OCCODE=DEAD}	0.21	0.22	1.03	0.93	96.11%
4	{DOAC=N,DTHROMB=N} => {OCCODE=DEAD}	0.21	0.22	1.02	0.95	96.67%
5	{DOAC=N} => {OCCODE=DEAD}	0.21	0.22	1.02	0.96	98.56%
6	{DAP=N,DHAEMD=N,DTHROMB=N} => {OCCODE=DEAD}	0.20	0.22	1.01	0.91	92.78%
7	{DAP=N,DHAEMD=N} => {OCCODE=DEAD}	0.20	0.22	1.01	0.93	94.67%
8	{DAP=N,DTHROMB=N} => {OCCODE=DEAD}	0.21	0.22	1.01	0.94	95.67%
9	{DAP=N} => {OCCODE=DEAD}	0.21	0.22	1.01	0.96	97.56%
10	{DTHROMB=N} => {OCCODE=DEAD}	0.21	0.22	1.00	0.98	98.00%
11	{DIVH=N,DTHROMB=N} => {OCCODE=DEPENDENT}	0.39	0.41	1.00	0.96	96.22%

No	RULE	Support	Confidence	Lift	Coverage	Akurasi
12	{DHAEMD=N,DMAJNCH=N} => {OCCODE=DEPENDENT}	0.39	0.41	1.00	0.96	96.40%
13	{DTHROMB=N,DMAJNCH=N} => {OCCODE=DEPENDENT}	0.40	0.41	1.00	0.98	97.06%
14	{DTHROMB=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=DEPENDENT}	0.39	0.51	1.26	0.76	94.95%
15	{DTHROMB=N,DMAJNCH=N,DCAUSE=unknown,CMPLHEP=Y} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.72	95.40%
16	{DTHROMB=N,DCAUSE=unknown,CMPLHEP=Y} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.73	95.68%
17	{DTHROMB=N,DCAUSE=unknown,CMPLASP=Y} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.73	95.83%
18	{DSCH=N,DTHROMB=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.74	96.69%
19	{DSCH=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.74	96.83%
20	{DSCH=N,DTHROMB=N,DCAUSE=unknown} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.75	96.98%
21	{DMAJNCH=N,DCAUSE=unknown,CMPLHEP=Y} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.74	96.69%
22	{DCAUSE=unknown,CMPLHEP=Y} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.74	96.98%

No	RULE	Support	Confidence	Lift	Coverage	Akurasi
23	{DSCH=N,DCAUSE=unknown} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.75	97.12%
24	{DMAJNCH=N,DCAUSE=unknown,CMPLASP=Y} => {OCCODE=RECOVER}	0.16	0.22	1.30	0.74	97.12%
25	{DMAJNCH=N,DCAUSE=unknown,CMPLASP=Y} => {OCCODE=NRECOVER}	0.19	0.26	1.30	0.74	95.73%
26	{DCAUSE=unknown,CMPLASP=Y} => {OCCODE=NRECOVER}	0.19	0.26	1.29	0.75	96.11%
27	{DSCH=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	0.19	0.26	1.29	0.74	96.49%
28	{DIVH=N,DMAJNCH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	0.19	0.26	1.29	0.75	96.24%
29	{DSCH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	0.19	0.25	1.29	0.75	96.74%
30	{DMAJNCH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	0.20	0.25	1.29	0.77	99.62%
31	{DIVH=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	0.19	0.25	1.29	0.75	96.49%
32	{DOAC=N,DCAUSE=unknown} => {OCCODE=NRECOVER}	0.19	0.25	1.28	0.75	95.23%
33	{DCAUSE=unknown} => {OCCODE=NRECOVER}	0.20	0.25	1.28	0.78	100.00%
34	{DMAJNCH=N,CMPLASP=Y} => {OCCODE=NRECOVER}	0.19	0.20	1.02	0.95	95.73%

No	RULE	Support	Confidence	Lift	Coverage	Akurasi
35	{CMPLASP=Y} => {OCCODE=NRECOVER}	0.19	0.20	1.01	0.95	96.11%
36	{DSCH=N,DMAJNCH=N} => {OCCODE=NRECOVER}	0.19	0.20	1.01	0.95	96.49%
37	{DIVH=N,DMAJNCH=N} => {OCCODE=NRECOVER}	0.19	0.20	1.01	0.96	96.24%
38	{DMAJNCH=N} => {OCCODE=NRECOVER}	0.20	0.20	1.00	0.99	99.62%
39	{DSCH=N} => {OCCODE=NRECOVER}	0.19	0.20	1.00	0.96	96.74%
40	{DIVH=N} => {OCCODE=NRECOVER}	0.19	0.20	1.00	0.96	96.49%
	Rata - rata	0.21	0.25	1.15	0.85	96.44%

RIWAYAT PENULIS



Erina Siska Dewi dilahirkan di Malang pada tanggal 3 November 1991. Gadis yang disapa Siska ini menghabiskan waktu kecilnya di desa Poncokusumo Kabupaten Malang. Menempuh pendidikan dari TK hingga SMP di Poncokusumo, yakni MI Sunan Muria Poncokusumo, dan SMPN 1 Poncokusumo. Jenjang pendidikan berikutnya Siska memilih untuk menempuh di SMK Telkom Sandhy

Putra Malang dan lulus pada tahun 2009. Setelah lulus penulis tak langsung melanjutkan pendidikan ke jenjang kuliah, dia memilih untuk bekerja selama satu tahun. Penulis mendapatkan kesempatan untuk melanjutkan ke jenjang perguruan tinggi setelah diterima sebagai salah satu mahasiswa Sistem Informasi ITS pada tahun 2010.

Aktivitasnya selain kuliah cukup beragam dengan mengikuti keanggotaan organisasi mahasiswa ITS dan beberapa kepanitiaan dari berbagai kegiatan di ITS. Pada tahun keduanya penulis sempat menjadi staff di dua organisasi yang berbeda, yakni departemen Dalam Negeri HMSI ITS dan kementerian Sosial Masyarakat BEM ITS. Selain itu, selama menyandang gelar sebagai mahasiswa, penulis menyempatkan diri menjadi asisten praktikum beberapa matakuliah. Dan akhirnya memilih Sistem Pendukung Keputusan sebagai bidang minatnya dalam pengerjaan tugas akhir karena ketertarikannya dalam *Data mining*.